PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C07K 14/52, A61K 38/19

A2

(11) International Publication Number: **WO 96/34887**

(43) International Publication Date: 7 November 1996 (07.11.96)

(21) International Application Number: PCT/GB96/01082

(22) International Filing Date: 7 May 1996 (07.05.96)

(30) Priority Data:

9509263.1	5 May 1995 (05.05.95)	GB
9607505.6	11 April 1996 (11.04.96)	GB

(71) Applicants *(for all designated States except US)*: IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY & MEDICINE [GB/GB]; Sherfield Building, Imperial College, London SW7 2AZ (GB). LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE [GB/GB]; Keppel Street, London WC1E 7HT (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MILLER, Andrew, David [GB/GB]; 27 Fairlawn Grove, Chiswick, London W4 5EJ (GB). RAYNES, John, Graham [GB/GB]; 112 St Albans Road, Sandridge, St Albans AL4 9LJ (GB).

(74) Agents: CHAPMAN, Paul, William et al.; Kilburn & Strode, 30 John Street, London WC1N 2DD (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: ANTISENSE PEPTIDES

(57) Abstract

Antisense peptides are provided which are useful in antagonising the biological effects of target molecules. In particular, the target molecule is a cytokine, e.g. IL-1α or IL-1β, TNFα or IL-8 and the antisense peptides thus find use in treating or preventing conditions mediated by these cytokines, for instance inflammatory conditions or cancer.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

ANTISENSE PEPTIDES

The present invention relates to antisense peptides or polypeptides capable of binding to a target ligand. In particular, it relates to antisense peptides which will bind to biologically active molecules such as IL-1α and/or IL-1β, TNF and Eotaxin, thereby altering their biological activity, and their use in medicine.

The reason why DNA is an anti-parallel double helix has traditionally been thought to derive from the requirement for replication of the genetic code. The sense strand provides the code and the antisense the means of propagating that code. However, another relationship is now thought to exist between sense and complementary /antisense strands which suggests a more substantive role for antisense DNA. Blacock and Smith (*Bioch.Biophys. Res.Comm.*, **121**:203 (1984)) observed that antisense DNA was able to code for peptides which are essentially the hydropathic complement of those dictated by sense DNA. This observation was followed by the subsequent discovery that sense peptides and their hydropathically complementary antisense peptides are able to interact (Bost *et al*, *PNAS USA*, **82**:1372 (1985)).

This later observation has now been elaborated into a Molecular Recognition Theory (MRT) to account for the interactions of protein ligands with their receptor targets and which has been used to explain a variety of interactions including IL-2 with its receptor (Weigent *et al*, *Bioch.Biophys.Res.Comm.*, **139**:367 (1986); Bost *et al*, *Bioch.Biophys.Res.Comm.*, **128**:1373 (1985)), collagen with collagenase (de Souza and Brentani *J.Biol.Chem.*, **267**:13763 (1992)) and cystatin C with C4 (Ghiso *et al*,

PNAS USA, **87**:1288 (1990)).

Furthermore, a recently described method for the affinity purification of IL-1β from complex synthetic mixtures and cell lysates (Fassina *et al*, *Int.J.Peptide Protein Res.*, **39**:549 (1992); Fassina & Cassani, *Biochem.J.*, **282**:773 (1992)) employed a polypeptide designed to act as a complement to a specified region of IL-1 sequence to bind the protein with high affinity, behaviour typically associated with receptors and antibodies. In addition, antisense peptides have been used as epitopes to generate antiidiotypic antibodies (Bost & Blalock *Meth.Enzymol.*, **178**:63 (1989); Araga *et al*, *PNAS USA*, **90**:8747 (1993)).

In general, hydrophobic amino acids are complemented by hydrophilic amino acids and vice versa, while, on average, "uncharged" residues are generally complemented by similar residues. This pattern was shown to represent the origin of the protein-receptor binding pair interactions for the related protein, interleukin-2 (Kuo & Robb *J.Immunol.*, **137**:1538-1543 (1986)), and it has been further postulated to result in the relationships that typify the immune system. The early work to test the validity of this pattern in natural assay systems was based on high-affinity binding of complements to the naturally occurring peptides ACTH (adrenocorticotrophic hormone) and γ-endorphin (Bost *et al*, (1985) *supra*). However, to date, much of the work has centred on the development of antibody-like molecules and affinity reagents to probe protein-receptor interaction.

Thus, previous research with the use of such anti-sense techniques has centred on their value as affinity agents and probes to purify target molecules from complex

mixtures.

We have now found that anti-sense peptides can be designed which exert an antagonistic effect on target ligands by virtue of their ability to bind to complementary "sense" peptide sequences contained therein.

This novel approach will allow the design of novel, potentially effective, therapeutic agents for a whole range of conditions mediated by polypeptides or proteins. Examples of suitable targets are discussed below.

The term interleukin-1 (IL-1) encompasses the two structurally similar and potent cytokines, interleukin-1α and -1β, which both have a MW of 17kDa, and almost identical tertiary structures despite being the product of two separate genes (Clore *et al*, *Biochemistry*, **29**:5671-5676 (1991); Clore *et al*, *Biochemistry*, **30**:2315 (1991); Driscoll *et al*, *Biochemistry*, **29**:3542-3556 (1990); Finzel *et al*, *J.Mol.Biol.*, **209**:779-791 (1989); Graves *et al*, *Biochemistry*, **29**:2679-2684 (1990); Priestle *et al*, *EMBO J.*, **7**:339-343 (1988); Priestle *et al*, *Cytokines.Lipocortins*, **349**:297-307 (1990). The two forms act, with varying affinities, through the same cell surface receptors (IL-1 RI and IL-1 RII and accessory chain, Dower *et al*, *Nature*, **324**:266-268 (1986); Killian *et al*, *J.Immunol.*, **136**:4509 (1986); Sims & Dower, *J.Biol.Chem.*, **6**:112 (1990)) and, with few exceptions, elicit similar responses (Boraschi *et al*, *Eur.J.Immunol.*, **20**:317 (1990); Calkins *et al*, *Biochem.Biophys.Res.Commun.*, **167**:548-553 (1990)).

Through widespread effects, these proteins perform central mediatory roles in immunity, haematopoiesis and the inflammatory response as well as in rheumatoid

arthritis and septic shock (Dinarello, *Blood*, **77**:1627-1652 (1991); Dinarello and Wolff, *N.Eng.J.Med.*, **328**:106 (1993)). Both have been implicated in the amyloid pathology of head trauma associated with Alzheimer's disease (AD) and other neurodegenerative disorders, apparently upregulating the Amyloid Precursor Proteins (APP) (Vandenabeele & Friers, *Immunol.Today*, **12**:217 (1991); Buxbaum *et al*, *PNAS USA*, **89**:10075 (1992); Vasilakos *et al*, *FEBS Lett.*, **354**:289 (1994)), a ubiquitous family of transmembrane glycoproteins expressed throughout the body tissues.

Tight control of IL-1 activity within biological systems is normally maintained by mechanisms which include a third member of the IL-1 family, namely the naturally occurring N-glycosylated interleukin 1 receptor antagonist (IL-1ra) (Eisenberg *et al*, *Nature*, **343**:341 (1990); Carter *et al*, *Nature*, **344**:633-638 (1990); Hannum *et al*, *Nature*, **343**:336-340 (1990); Seckinger *et al*, *J.Immunol.*, **139**:1541 (1987); Vigers *et al*, *J.Biol.Chem.*, **269**:12874 (1994); Stockman *et al*, *FEBS Lett.*, **349**:79 (1994)), which binds competitively to the IL-1 receptors, but, lacking the necessary trigger domain, fails to elicit any comparable biological response (Dripps *et al*, *J.Biol.Chem.*, **266**:10331-10336 (1991)).

In clinical trials, IL-1ra was found to be more effective than traditional treatments for reduction of joint inflammation and discomfort. However, due to its short half-life, and high dosage requirements, it is of little therapeutic value. Potential therapeutic strategies aimed at modulating IL-1 pathological activity have therefore been based on the development and introduction of more stable IL-1ra analogues; site-directed

mutagenesis studies have highlighted several mutations in the IL-1 sequence which give rise to partial antagonism (Ferreira *et al*, *Nature*, **334**:698-700 (1988)). Other recent attempts to inhibit IL-1 effects include soluble receptor antagonists and mimics (Bates *et al*, *Expert Opinion in Therapeutic Patents*, **4**:917 (1994)) as well as pyridinyl-imidazole inhibitors (Lee *et al*,1 *Nature*, **372**:739 (1994)).

TNFα is widely appreciated as a principal mediator of systemic responses to sepsis and injury (Beutler, B., and Cerami, A., *Ann. Rev. Biochem.*,**57**:505-518 (1988)). Produced by inflammatory cells in response to diverse infectious stimuli and tissue injury, TNFα induces a cascade of mediators that direct host immunological functions (Fong, *et al*, *J.Exp. Med.*, **170**:1627-1633 (1989)). While TNFα may thus serve as an essential messenger in host defense, the excessive tissue production of TNFα can mediate detrimental sytem effects by acutely precipitating a syndrome similiar to that of septic shock (Tracey, *et al*, Science **234**:470-474 (1986)), and lesser degrees of chronic TNFα production appear to induce anorexia and cachexia (Moldawer, *et al*, *Am. J. Physiol.*, **254**:G450-G456 (1988); Tracey, *et al*, *J.Exp. Med* **167**:1211-12278 (1988)). Thus, pathologic conditions may result from the excessive production and activity of TNFα.

The active form of TNFα is believed to be a homotrimer with 17-kDa subunit polypeptides (Smith., and Baglioni, C., *J. Biol. Chem.*, **262**:6951-6954 (1987)). TNFα and TNFβ, a related lymphokine, activities are mediated through two distinct receptors, TNFR-p55 and TNFR-p75 (Loetscher, *et al*, *Cell*, **61**:361-370 (1990)). Both receptors bind TNFα

and TNFβ with simililar affinities, but they are independently regulated.

Eotaxin was recently identified as an important eosinophil chemoattractant detected in bronchoalveolar lavage fluid (BAL) after allergen challenge of sensitised guinea-pigs (Jose *et al*, *J.Exp.Med*, **179**:881-887 (1994)). The potency of this chemokine has been demonstrated by low dose induction of eosinphil accumulation in guinea-pig airways and skin *in vivo* , and by comparable eosinophil trafficking in guinea-pig and human cells *in vitro* (Jose *et al* (1994) *supra*; Bousquet.J. and Charez.P., N., *Eng.J.Med*, **323**:1033-1039 (1994)). The role of eosinophil accumulation in Ig-E mediated allergic responses is well known, particularly for asthma (Warringa *et al*, *J.Allergy Clin.Immunol*, **91**:1198-1205 (1993)), eczema (Leiferman.K.M., *J.Am.Acad.Derm.*, **24**:1101-1112 (1993)), rhinitis (Cantani *et al*, *J.Invest.Allerg.Clin.Immunol*, **2**:181-186 1992)) and various parasitic infections (Gounni *et al*, *Am. Rev.Respir.Dis.*, **131**:373-376 1985)). Eotaxin appears to be unique, so far, among the chemokines in its ability to selectively activate eosinophils which accumulate in both guinea-pig skin and airways and is thus a selective subject for inhibition studies.

Interleukin-8 (IL-8) is a CXC class chemokine structurally related to platelet factor 4 (Baggiolini.M., and Clark-Lewis.I., *FEBS*, **307**:97-101 (192)). It is produced by phagocytes and mesenchymal cells exposed to inflammatory stimuli and activates neutrophils inducing chemotaxis, exocytosis and the respiratory burst (Seitz *et al*, *J.Clin. Inv.*, **87**:463 (1991)). As a product of different types of cells it can arise in any tissue when

the levels of IL-1 and TNF are enhanced.

Structure-activity relation studies indicate that IL8 binds at the N-terminus (Clark-Lewis *et al*, J.Biol.Chem, **266**:23128 (1991)). No receptor binding or neutrophil activation was observed when the N-terminal sequence Glu-Leu-Arg (ELR) that precedes the first cysteine is deleted. The role of this motif as the main structural determinant of receptor binding is confirmed by site directed mutagenesis studies in which IL-8 mutants, containing alanines in place of the ELR sequence, are all chemotacticly inactive (Hebert *et al*, *J.Biol.Chem.* **266**:18989 (1991)).

In a first aspect, therefore, the present invention provides a peptide or polypeptide comprising an amino acid sequence which is antisense to a target peptide or polypeptide sequence, wherein said antisense peptide or polypeptide binds to the target peptide or polypeptide, thereby altering the biological activity of the target peptide or polypeptide or the biological activity of a target molecule which comprises the target peptide or polypeptide.

The term "anti-sense" has heretofore generally been applied to nucleic acid sequences which are capable of binding to complementary nucleic sequences. For example, anti-sense DNA sequences can be generated which in turn can generate mRNA sequences which will bind to mRNA produced from coding/sense strands of DNA, thereby preventing translation. In the context of the present invention, the term "anti-sense peptide or polypeptide" refers to a peptide or polypeptide coded for by a nucleic acid sequence complementary to the nucleic acid sequence coding for the target sequence. An anti-sense peptide or

polypeptide within the context of the present invention also includes a peptide or polypeptide, at least part of whose sequence is anti-sense to a target sequence.

Clearly, as the skilled man will appreciate, the concept of targeting a peptide or polypeptide sequence with an antisense peptide or polypeptide can be applied to relatively small biologically active peptides in order to affect their biological activity. Alternatively, the target sequence will form part of a larger molecule, with the target sequence being involved in the biological activity of the molecule.

In one embodiment of this aspect of the invention, the anti-sense peptide or polypeptide can act as an antagonist to or inhibitor of the biological activity of the target sequence or molecule.

A preferred group of target molecules are cytokines, for example IL-1α and/or IL-1β, IL-8 or TNFα. In the case of IL-1α or IL-1β the anti-sense peptide is preferably anti-sense to a target sequence located within the region of residues 47-55 of IL-1β. Examples of such antisense peptides include:-

N-VITFFSL; and
N-VITFFS.

In the case of TNFα the antisense peptide is preferably antisense to a target sequence located within the region of residues 83-91 or 29-34 of TNFα. Examples of suitable peptides include:-

N-DLGLVRDGD;
N-LGLVRDG; and

N-IGPAVQ.

In the case of IL-8 an example of a suitable antisense peptide is:-

N-SKLFS.

Another example of a suitable target molecule for the antisense approach is Eotaxin. Preferably, the antisense peptide is antisense to a target sequence located within the region 43-49 or 45-50 of Eotaxin. Examples of suitable antisense peptides include:-

N-DILGQFG; and
N-HFVRFD.

As discussed above, these antisense peptides could form part of a larger peptide or polypeptide. The key property which any such larger sequence must possess is of course the ability to bind to the target sequence.

The skilled man will appreciate that the concept of "anti-sense" peptides or polypeptides can be applied generally to alter the biological properties of a range or "targets". Techniques are readily available, as discussed in the examples below, for identification of target sequences, which may form part of larger, biologically active, molecules. Once such target sequences have been identified "antisense" peptides or polypeptides can be generated (again using standard techniques) and tested against the target molecule.

Thus, in a second aspect, the invention provides an anti-sense peptide or polypeptide for use in altering the biological activity of a target sequence or molecule.

In particular, anti-sense peptides or polypeptides would be useful as therapeutic agents by virtue of their ability to alter biological activity of a target sequence or molecule.

In a further aspect, therefore, the invention provides an anti-sense peptide or polypeptide for use in medicine, particularly in the treatment or prevention of an inflammatory condition and/or cancer.

Generally, the anti-sense peptides or polypeptides of the invention will find application in medicine in the form of a pharmaceutical formulation. In a fourth aspect, therefore, the invention provides the use of an antisense peptide or polypetide as defined herein in the manufacture of a medicament for use in the prophylaxis or treatment of a condition mediated by a cytokine.

In a fifth aspect the invention provides a pharmaceutical formulation comprising at least one anti-sense peptide or polypeptide as defined herein, together with one or more pharmaceutically acceptable carriers, diluents or excipients.

In particular embodiments, the target molecule is IL-1α and/or IL-1β, TNFα, Eotaxin or IL-8 and the anti-sense peptides or polypeptides of the invention can be used to inhibit the biological action of these target molecules and thus find use in treating inflammatory conditions, e.g. septic shock, rheumatic diseases and degenerative arthropathies, including rheumatoid arthritis as well as, in the case of TNFα, cancer.

In a sixth aspect, the present invention provides a

method of treating a condition mediated by a cytokine which comprises administering to a patient an effective amount of an antisense peptide or polypeptide as defined herein

In a final aspect the invention provides a method for the prophylaxis or treatment of an inflammatory condition which comprises administering to a patient an effective amount of an antisense peptide or polypeptide as defined herein, preferably in the form of a pharmaceutical formulation.

Preferred features of each aspect of the invention are applicable to each other aspect *mutatis mutandis*.

The invention will now be described by way of the following examples, which are not to be construed as in any way limiting the invention.

The examples refer to the figures in which:-

FIGURE 1a: shows an alignment of the gene-derived amino acid sequences of IL-1α, IL-1β and IL-1ra;

FIGURE 1b: shows the DNA sequence coding for the β-bulge of IL-1β (Boraschi loop) in alignment with the complementary/antisense DNA sequence;

FIGURE 1c: shows a comparison of the Kyte-Doolittle hydropathic profiles of the Boraschi loop sequence, QGEESND (■, antisense peptide, VITFFSL (□) and control peptide (●);

FIGURE 2a: shows the dose dependent inhibition by

antisense peptide VITFFSL of IL-1β stimulated synthesis of serum amyloid A (SAA) (black bars) and haptoglobin (hatched bars) in HuH7 hepatoma cell supernatants after 48h of stimulation;

FIGURE 2b: shows the inhibition of IL-1β interaction with soluble human IL-1 receptor type II (sIL-1 RII) by antisense peptides, VITFFSL (■) and VITFFS (□);

FIGURE 3a: shows surface plasmon resonance (SPR) affinity profiles obtained from the interaction of antisense peptide, VITFFSL, with immobilised IL-1β. Peptide concentrations were 20μM (—), 40μM (- -), 70μM (···), 90μM (-·-), 100μM (-···-) and 200μM (—);

FIGURE 3b: shows the maximum affinity profile response changes (after background correction) plotted as afunction of VITFFSL concentration. Results were obtained with immobilised IL-1β (surface concentration 22μg/ml) in the absence (△) and presence (O) of sense peptide QGEESND (equimolar with VITFFSL), and immobilised IL-1α (△) (surface concentration 15μg/ml);

FIGURE 4: shows a comparison of the hydropathic profiles of TNFα residues 83-91 and the corresponding antisense peptide;

FIGURE 5: shows a comparison of the hydropathic profiles of the antisense peptide for region 83-91 of TNFα and the region 91-99 of TNFR 55;

FIGURE 6: shows a comparison of the hydropathic

profiles of TNFα residues 83-91, TNFRp55 (91-99) and TNFRp75 (91-99);

FIGURE 7: shows the results of TNF inhibition assays using two antisense peptides;

FIGURE 8: shows a comparison of the average hydropathy of huMCP-1 and gp EOTXAIN;

FIGURE 9: shows a comparison of the average hydropathy of hEOTAXIN and hMCP-1;

FIGURE 10: shows a comparison of the hydropathic profiles of gp EOTAXIN (45-50), a corresponding antisense peptide (HFVRFD) and CCCKR3 receptor fragment 146-152;

FIGURE 11: shows a comparison of the hydropathic profiles of a portion of the sequence of IL-8 (AKELR), a corresponding antisense peptide (SKLFS) and IL-8R sequence (AKFLT).

EXAMPLE 1:IL-1

Standard Methods

Peptide generation

All peptides were generated to a standard Fmoc protocol using a fully automated Applied Biosystems 431A synthesizer, software version 1.1 using Fmoc protected amino acids acquired from Bachem (UK) Ltd and the Rink Amide MBHA resin purchased from Novabiochem.

Peptide purification.

Freeze dried peptides were desalted on P2 Biogel gel filtration column (2 cm x 30 cm) eluted with 0.1% TFA

then loaded onto a Pharmacia biotech. Pep RPC HR10/10 column and eluted at 1.5 ml/min on linear gradient from 0.1% $TFA_{(aq)}$ to 100% acetonitrile, 0.1% TFA, and their identity verified by +ve FABMS.

The first stage in the design antisense peptide inhibitors was to identify a suitable target region, from amongst the overlapping functional regions of IL-1α and IL-1β, against which antisense peptides could be designed to act. Therefore, the X-ray crystal structure of IL-1β (Finzel *et al*, (1989) *supra*; Priestle *et al*, (1988), *supra*), IL-1α (Graves *et al*, (1990), *supra*) and IL-1ra (Vigers *et al*, (1994), *supra*) were reviewed in conjunction with the results of recent mutational studies performed on all three proteins (Evans *et al*, *J.Biol.Chem.*, **270**:11477 (1995); Labriola-Tompkins *et al*, *Prot.Eng.*, **6**:535 (1993); Labriola-Tompkins *et al*, *PNAS USA*, **88**:11182 (1991); Grutter *et al*, *Prot.Eng.*, **7**:663 (1994); Gayle *et al*, *J.Biol.Chem.*, **268**:22105 (1993); Kawashima *et al*, *Prot.Eng.*, **5**:171 (1992)).

When an overlay of all three structures was made it was observed that both IL-1β and IL-1α possess a β-bulge structure, proximal to the known receptor binding amino acid residues, which is absent in the antagonist IL-1ra. This appeared to represent the only significant difference between the three-dimensional structures of IL-1ra and the other two IL-1 isoforms. This absence of secondary structure was also mirrored at the primary amino acid sequence level, where IL-1ra was found to have no equivalent stretch of amino acids to the β-bulge regions of IL-1β (residues 48-54[mature protein sequence]) or IL-1α (residues 60-66[mature protein sequence]) (fig 1a).

These structural differences suggested that the β-bulge regions of IL-1β and IL-1α might be promising targets for antisense peptide inhibitors. We were confirmed in this approach firstly by the suggestion (Auron *et al*, *Biochem.*, **31**:6632 (1992)) that these same β-bulge regions may act as the "early trigger" for IL-1 receptor mediated gene transcription, and secondly by reports from Boraschi and coworkers (Antoni *et al*, *J.Immunol.*, **137**:3201 (1986); Boraschi *et al*, *J.Exp.Med.*, **168**:675 (1988)) that a peptide, corresponding in sequence to the β-bulge region IL-1β (residues 47-55[mature protein]), possesses partial IL-1 agonist activity.

the β-bulge of IL-1β is more pronounced than that of IL-1α (fig 1a) and therefore it was anticipated that β-bulge directed inhibitors might be more effective against IL-1β than IL-1α. As a result, antisense peptides were designed to primarily target the β-bulge region of IL-1β, which was re-christened the Boraschi loop.

Antisense peptides to the Boraschi loop were designed with reference to the DNA sequence of IL-1β (March *et al*, *Nature*, **315**:641 (1985)). Having identified the DNA sequence coding for the loop, the complementary/antisense DNA sequence was deduced and the code translated in the 5'→3' direction (fig 1b). Two antisense peptides were then synthesised (see above) on the basis of the antisense code. The first with the sequence VITFFS, complementary to Boraschi-loop segment GEESND (IL-1β residues 49-54), and the second with the sequence VITFFSL, complementary to the Boraschi-loop segment QGEESND (IL-1β residues 48-54). Both peptides were hydrophobic and were therefore prepared as C-terminal amide derivatives to aid aqueous solubility. The Kyte-

Doolittle hydropathic profile of VITFFSL plotted (fig 1c) with the profile of QGEESND illustrates the mutual complementarity of their hydropathic profiles. The hydropathic profile of VITFFSL was also found as well to be reasonably complementary to the profile of the IL-1α β-bulge sequence, KSSKDDA (residues 60-66), leading us to believe that VITFFSL and VITFFS may also target IL-1α, even though these peptides had not been primarily designed to do this. Therefore tests were carried out to try to inhibit the effects of both IL-1α and IL-1β.

The antisense peptides were tested for biological effect using an HuH7 hepatoma cell line assay system (Bevan & Raynes, *J.Immunol*, **147**:2574 (1991)). In this assay system serum amyloid A (SAA) and haptoglobin are induced directly in response to IL-1 (Raynes *et al*, *Clin.Exp.Immunol.*, **83**:448 (1991)). Antisense peptide was predissolved in DMSO (10mg/ml) and diluted, to various concentrations (see fig 2a), in the wells of 24-well plates which contained confluent HuH7 cells under the stimulation of IL-1β(1ng/ml). Specific protein concentrations were measured by ELISA.

Antisense peptides, VITFFSL and VITFFS, were found to inhibit both IL-1β and IL-1α stimulated synthesis of SAA and haptoglobin in a dose dependent manner (fig2a; table 1). The levels of inhibition are approaching or even exceeding those observed when either IL-1ra or soluble human IL-1 type II receptor (sIL-1 RII) were used as inhibitors in the same assy system (table 1). The data show (fig 2a: table 1) that SAA was inhibited more readily than haptoglobin consistent with previous observations with IL-1ra (Bevan & Raynes (1991), *supra*).

Table 1

PEPTIDE	%SAA inhibition		% Haptoglobin inhibition	
	IL-1α*	IL-1β*	IL-1α*	IL-1β*
N-VITFFSL'	75	78	ND	66
N-VITFFS'	55	65	ND	57
IL-1ra'	100	100	100	100
sIL-1 RII'	25	90	38	75
N-VFITSFL'	<10	<10	<10	<10
N-LSFFTIV'	<10	<10	<10	<10
N-LLSLLPV'	<10	<10	<10	<10
N-LLSLLRV'	<10	<10	<10	<10

* 1ng/ml
' 20μg/ml
' 10μg/ml
' 10μg/ml (Bevan & Raynes, (1991), *supra*)

In order to test the specificity of IL-1 inhibition by the two antisense peptides, four more peptides were synthesised as controls and tested in the assay (table 1). The choice of controls was dictated by the apparent importance of hydropathic complementarity in determining physical interaction between sense and antisense peptides (Shai *et al*, *Biochem*, **28**:8804 (1989); Shai *et al*,

Biochem, **26**:669 (1987)). This suggested that reordering the amino acid sequence of the antisense peptide, to alter the hydropathic profile, should abolish the interaction with the sense peptide, whilst by contrast, a peptide of similar hydropathic profile to the antisense peptide, but different amino acid sequence, would be expected to interact with the original sense peptide.

In keeping with the first part of this analysis, VFITSFL (a reordered peptide with altered hydropathic profile; fig 1d) failed to measurably inhibit IL-1. However, LSFFTIV (the reverse peptide with an identical profile) and LLSLLRV (a peptide with similar profile but different sequence) also failed to inhibit IL-1 in contradiction to the second part of the analysis. Finally, there was some precedent (Bost *et al* (1985), *supra*) to indicate that a peptide whose sequence was derived by translating the antisense codon in the 3'→5' direction (as opposed to the 5'→3' direction used above) should interact with a sense peptide. However, the corresponding peptide LLSLLPV was also found not to inhibit IL-1. Therefore, the inhibition of IL-1 by VITFFSL and VITFFS appeared very specific.

The origin of this specificity was analysed with the aid of a surface plasmon resonance (SPR) biosensor which was used to determine if antisense peptide VITFFSL was directly binding to IL-1β and IL-1α. Both cytokines were immobilised on SPR cuvettes and treated with increasing concentrations of VITFFSL under conditions comparable to the HuH7 assay. In both cases, a significant binding interaction was observed (fig 3). Analysis of the affinity profiles revealed that dissociation constants (K_d) (table 2) for the interaction with both cytokines were comparable to those measured for the interaction of

other sense and antisense peptides (Shai *et al*, (1989), *supra*). Moreover, the stoichiometry of association was calculated to be 4-5 peptides per immobilised cytokine molecule, a figure which also agrees with previous peptide/antisense peptide studies (Shai *et al*, (1989), *supra*). In order to investigate the specificity of the binding interaction, six alternative proteins (IL-1ra, interleukin-4 (IL-4), interleukin-5 (IL-5), interleukin-8 (IL-8), lysozyme and the glycoprotein CD59) were immobilised on SPR cuvettes and treated with VITFFSL in the same way. VITFFSL did not interact significantly with any of these six control proteins suggesting that the binding interaction of VITFFSL with IL-1β nd IL-1α was specific.

Table 2

Peptide		k_{ass} ($M^{-1}s^{-1}$)	10^4x K_{diss} (s^{-1})	K_d (μM)
VITFFSL	IL-1α	56.3	6.5	11.4
	IL-1β	64.5	6.6	10.2
LSFFTIV	IL-1α	88.9	249.9	281.1
	IL-1β	71.1	113.7	160.0
VFITSFL	IL-1α	*	*	*
	IL-1β	*	*	*
LLSLLRV	IL-1α	60.8	210.8	345.6
	IL-1β	86.5	252.4	291.3
LLSLLPV	IL-1α	52.4	318.5	607.7
	IL-1β	48.4	203.8	419.9

* Too weak to be determined

The interaction of the four control peptides (VFITSFL, LSFFTIV, LLSLLRV and LLSLLPV) with IL-1β and IL-1α was also investigated with the SPR biosensor. The data obtained (table 2) appears to offer an explanation for the specificity of peptides VITFFSL and VITFFS as IL-1 inhibitors. In keeping with the importance of hydropathic complementarity in antisense/sense peptide interactions, the reordered peptide VFITSFL was found not to interact with either IL-1β or IL-1α. The other three peptides, whose hydropathic profiles are similar to VITFFSL, did interact with both cytokines but at least an order of magnitude more weakly. The association rates, k_{ass}, of these three peptides and VITFFSL with the cytokines were the same within experimental error. Therefore, the tighter binding shown by VITFFSL resulted from the slower dissociation rate of this peptide. Presumably, the failure of peptides LSFFTIV, LLSLLRV and LLSLLPV to show inhibition of IL-1 within the experimental limits of the HuH7 assay (<10% inhibition) is a reflection of their weaker associations (between 20 and 60-fold worse) with both IL-1β and IL-1α.

Having established a specific association between VITFFSL and both IL-1β and IL-1α, a receptor binding assay was then carried out to establish if VITFFSL and VITFFS could also inhibit the interaction of IL-1 with receptor. Accordingly, immobilised IL-1β was treated with soluble receptor sIL-1 RII in the presence and absence of both peptides (fig 2b). Il-1β was immobilised (4°C, 16h) on Immulon II plates in PBS buffer. The plates were blocked with PBS tween-20 (0.05%, v/v) containing BSA (1%, w/v) followed by the addition of sIL-1 RII (4ng/ml) in the presence of peptide at various concentrations (see fig 2b). In the event, both peptides were able to inhibit

receptor binding, the former by at least 63% compared to control binding in the absence of peptide.

The results of all the experiments described above, interlock to support the view that antisense peptides VITFFSL and VITFFS are specific inhibitors of IL-1 stimulated protein synthesis whose mechanism of action probably involves direct association with the cytokine thereby blocking receptor binding. Therefore, both peptides do appear to be acting as "mini receptor" inhibitors. The evidence suggests that the peptides are binding to the β-bulge structures which the cytokines possess. Firstly, the interaction of VITFFSL with IL-1βwas severely reduced in the presence of the sense peptide QGEESND, presumably because the sense peptide is competitive with IL-1β. Secondly, VITFFSL was unable to interact with IL-1ra (see above) which lacks the β-bulge structure (fig 1). Finally, both VITFFSL and VITFFS were found to be weak inhibitors of the association between IL-1β and the low affinity antibody BhrD2 which is specific to IL-1β amino acid residues 45-87 (incorporating the Boraschi loop structure).

EXAMPLE 2:TNFα

Standard Methods

Peptide generation

Peptides were generated to a standard Fmoc protocol using (i) a fully automated Applied Biosystems 431A synthesizer, software version 1.1 using Fmoc protected amino acids acquired from Bachem (UK) Ltd. and Rink Amide MBHA resin purchased from Novabiochem, and (ii) Schimadzu RF SPPS automated synthesizer.

Peptide purification

Freeze dried peptides were desalted on a P2 Biogel gel filtration column (2cm x 30cm) eluted with 0.1% TFA, 20% Acetonitrile (ACN) peptides were further purified by loading onto a Pharmacia biotech. Pep RPC HR 10/10 column and eluted at 1.0 ml/min on a linear gradient from 0.1%TFA, 20% ACN to 0.1%TFA, 100% ACN. Their identity is verified by +ve FABMS.

Testing on Iasys system

Test and control proteins were immobilised onto carboxymethyl dextran (CMD) coated cuvette surfaces at 37°C and pH 7.4, in PBS buffer using a standard NHS/EDC coupling protocol (described in the Iasys manual). Coupling times of 30-40 mins were allowed to generate a surface concentration suitable for analyte assay runs.

High pressure affinity chromatography(HPAC)

Peptides were attached to a Pharmacia activated CH sepharose 4B employing standard NHS coupling methods and packed on a CR10/10 column. Ligand binding is assessed by both zonal and continuous elution methods (Shai *et al*, *Biochemistry*, **26**:669-675 (1987)).

Site directed mutagenesis studies on TNFα have been carried out with a view to define the receptor binding site structural-functional relationship (Zhang *et al*, *J. Biol. Chem.*, **267**:24069-24075 (1992)). It appears that trimer formation is necessary for receptor binding and thus cytotoxic activity. Several TNFα surface mutants are found to allow trimer formation but impair cytotoxic activity, presumably through a lack of receptor activation. From these studies, four regions of TNFα appear critical for *in vitro* receptor biological

activity. Of these, surface residues 83-91 represent the most solvent exposed region, as viewed from the X-ray crystal structure with a Quanta Molecular Graphics program. A single point mutation of Tyr^{87} to any other residue identity impairs all biological activity. On this basis, antisense peptides (aFNT I and II) were designed to bind to this section of protein as shown below.

TNFα seq.83-91

N	Ile	Ala	Val	Ser	Tyr	Gln	Thr	Lys	Val	C
	I	A	V	S	Y	Q	T	K	V	
	A T C	G C C	G T C	T C G	T A C	C A G	A C C	A A G	G T C	
	T A G	C G G	C A G	A G C	A T G	G T C	T G G	T T C	C A G	
	D	G	D	R	V	L	G	L	D	
C	Asp	Gly	Asp	Arg	Val	Leu	Gly	Leu	Asp	N

aFNT I NH_2 D L G L V R D G D COOH

aFNT II NH_2 L G L V R D G COOH

Another of the TNFα solvent exposed active regions (29-34) is known to be critical in binding to the p75 receptor only. A complementary peptide to this region could specifically prevent p75 activation and thus allow the local cytotoxic TNF response mediated by the TNFRp55. This could provide the therapeutic basis for administering high doses of TNFα in anti-cancer therapy without the systemic toxicity dependent on p75 activation (Van Ostade *et al*, *Nature*, **361**:266-269 (1993)). An antisense peptide designed to bind to this region is shown below.

TNFα seq. 29-34

N	Leu	Asn	Arg	Arg	Ala	Asn	C
	L	N	R	R	A	N	
	C T G	A A C	C G C	C G G	G C C	A A T	
	G A C	T T G	G C G	G C C	C G G	T T A	
	Q	V	A	P	G	I	
C	Gln	Val	Ala	Pro	Gly	Ile	N

aFNT III NH_2 I G P A V Q COOH

The Molecular Recognition theory (MRT) theory purports an idea that antisense relationships form the basis of receptor-effector recognition. This supposition was tested theoretically for TNFα : if recognition between receptors and effectors is based on antisense, then an antisense peptide designed against a known activating region of TNFα might share some amino acid homology with conserved sections of the p55 and p75 receptors. An antisense homology search for TNF 83-91 revealed this to be the case: a 5 residue section of the nine residue effector sequence shared homology (including conservative substitutions) with a section of the p55 and p75 receptors. Moreover, the hydropathy plots for these receptor segments and for the antisense peptides aFNT were very similar and display a strong inverse correlation with the hydropathy trace for TNF 83-91 (figs 4,5,6).

Peptides were assayed for TNFα inhibitory activity using the L929 cytotoxicity assay. L929 cells are trypsinized, washed and resuspended at 8 x 10^5 cells/ml and 50μl added to wells of a 96 well plate. TNFα standards of 30, 10, 3, 1 0.3 and 0.1 U/ml were added and either 7.5 or 3.75 U/ml of TNF incubated with various concentrations of peptide for less than 30 min. before adding to the cells.

Actinomycin D was added to a final concentration of 1μg/ml to increase sensitivity. The plates were incubated for 24h, MTT was added and left for 4h before removal of the supernatant and colour determination, by dissolving in isopropanol:1-propanol (1:9 vol; 100μl) and measuring at 550nm. The results are shown in figure 7.

EXAMPLE 3:EOTAXIN

Standard methods were as for example 2.

The cDNA cloning and expression of eotaxin revealed interesting relationships between sequence homology and respective selectivities of other CC chemokines. It shared the greatest homology (53%) with human monocyte chemoattractant protein 1 (MCP1), a monocyte chemotaxant but not an eosinophil attractant in guinea-pig or human assay (Jose *et al*, (1994) *supra*). Less homolgy is observed for the human macrophage inflammatory protein (hMIP-1α, 31%) and hRANTES (26%) which are both inactive to eosinophils in guinea-pig skin but are active in human *in vitro* studies (Jose *et al*, (1994) *supra*). Based on previous experiments showing that complementary peptides can bind to one another and that this interaction is associated with an inverse correlation in hydropathy plot (according to the Kyte and Doolittle scale (Fassina *et al*, *Int.J.Peptide.Res*, **39**:549-556 (1992)), it was theorised that a comparison of both hydropathy plots and structural homology might yield some information on possible effector regions of these molecules. The analysis of average hydropathy plots for MCP1, gpEotaxin and huEotaxin revealed a striking similarity apart from three distinct regions (fig.8,9) . One of these regions on MCP1 correlated to a very solvent exposed 'loop' as viewed from the X-ray crystal structure.

Homology models of both gp and hu Eotaxin were constructed based on the Brookhaven Protein Databank file coordinates of MCP1 and the disparate region of hydropathic profile corresponded with the same solvent exposed loop (fig. 8). As MCP1 lacks eosinophil stimulatory activity in both guinea-pig and human assays, despite showing high sequence and structural homology to eotaxin, it was hypothesised that this loop sequence (45-49) in both hu and gp eotaxin are possible activation regions and further that their complementary peptides (see below) might bind to them and thus inhibit their biological activity.

hu EOTAXIN 43-49 and antisense sequence derivation

sense sequence NH_2	**Thr**	**Lys**	**Leu**	**Ala**	**Lys**	**Asp**	**Ile**	**COOH**
cDNA sequence	A C C	A A A	C T G	G C C	A A G	G A T	A T C	
	T G G	T T T	G A C	C G G	T T C	C T A	T A G	
	G	F	Q	G	L	I	D	
antisense seq. **COOH**	**Gly**	**Phe**	**Gln**	**Gly**	**Leu**	**Ile**	**Asp**	**NH_2**

NH_2 I L G Q F G COOH antisense to hu EOTAXIN 43-48

NH_2 D I L G Q F COOH antisense to hu EOTAXIN 44-49

gp EOTAXIN 45-50 and antisense sequence derivation

sense sequence NH_2	**Ile**	**Lys**	**Pro**	**Asp**	**Lys**	**Met**	**COOH**
	I	K	P	D	k	M	
cDNA sequence	A T C	A A A	C C T	G A C	C A A	A T G	
	T A G	T T T	G G A	C T G	G T T	T A C	
	D	F	R	V	F	H	
antisense seq. **COOH**	**Asp**	**Phe**	**Arg**	**Val**	**Phe**	**His**	**NH_2**

NH_2 H F V R F D COOH antisense to gp EOTAXIN 45-50

The receptors through which many of the CC chemokines

elicit responses have been cloned and expressed. Through cross-desensitization experiments, it was proposed that eosinophils have, among others, a shared receptor for eotaxin, CC CKR3 (Combadiere *et al*, *J.Biol.Chem.*, **270**;27:16941-16949 (1995)). Based on the concept that inverse hydropathy charcterised possible binding complements, it seems reasonable that an eotaxin receptor would include a region antisense to the putative effector region on eotaxin. This indeed was the case: a sequence of four residues were directly antisense to the putative activating loop on gp Eotaxin, whilst the two flanking residues were hydropathically complementary to the residues expected to be proximal when aligned N to C and C to N respectively (fig. 10 displays hydropathy plots and residue identity).

EXAMPLE 4:IL-8

Standard methods were as for example 2.

IL-8 in humans is known to elicit its response through two distinct seven transmembrane spanning receptors, IL-8R1 and IL-8R2. In accordance with the complementary approach outlined above, it was theorised that if a region of one of the known receptors was homologous to the antisense peptide sequence of the IL-8 activating region, then a synthetic equivalent of that sequence may inhibit IL8 activity by binding to it. A search for homology between the ELR antisense complementary sequences and the receptor sequences revealed several segments satisfying the antisense combinations; of these one was integral to a 5 residue sequence whose hydropathy profile displays excellent negative correlation to the AKELR IL-8 N terminal region (fig. 11).

CLAIMS

1. A peptide or polypeptide comprising an amino acid sequence which is antisense to a target peptide or polypeptide sequence, wherein said antisense peptide or polypeptide binds to the target peptide or polypeptide, thereby altering the biological activity of the target peptide or polypeptide or the biological activity of a target molecule which comprises the target peptide or polypeptide.

2. An antisense peptide or polypeptide as claimed in claim 1, which is wholly antisense for a target sequence.

3. An antisense peptide or polypeptide as claimed in claim 1 or claim 2, which binds to a target sequence which forms part of a larger target molecule.

4. An antisense peptide or polypeptide as claimed in any one of claims 1 to 3 which acts as an antagonist for or inhibitor of the target sequence or molecule.

5. An antisense peptide or polypeptide as claimed in any one of claims 1 to 4, wherein the target molecule is a cytokine.

6. An antisense peptide or polypeptide as claimed in claim 5, wherein the cytokine is IL-1α and/or IL-1β, TNFα or IL-8.

7. An antisense peptide or polypeptide as claimed in claim 6, which is antisense to a sequence within the region of residues 47-55 of IL-1β.

8. An antisense peptide or polypeptide as claimed in claim 7 which includes the amino acid sequence:

N-VITFFSL; or
N-VITFFS.

9. An antisense peptide or polypeptide as claimed in claim 6, which is antisense to a sequence within the region of residues 83-91 or 29-34 of TNFα.

10. An antisense peptide or polypeptide as claimed in claim 9 which includes the amino acid sequence:

N-DLGLVRDGD;
N-LGLVRDG; or
N-IGPAVQ.

11. An antisense peptide or polypeptide as claimed in claim 6 wherein the target molecule is IL-8 and the peptide or polypeptide includes the amino acid sequence:
N-SKLFS.

12. An antisense peptide or polypeptide as claimed in any one of claims 1 to 4 wherein the target molecule is EOTAXIN.

13. An antisense peptide or polypeptide as claimed in claim 12 which include the amino acid sequence:
N-DILGQFG; or
N-HFVRFD.

14. An antisense peptide or polypeptide for use in altering the biological activity of a target sequence or molecule.

15. An antisense peptide or polypeptide for use in medicine.

16. An antisense peptide or polypeptide as claimed in claim 14 or claim 15, modified by any one or more of the features of claims 2 to 13.

17. An antisense peptide or polypeptide as claimed in claim 16 for use in treating or preventing an inflammatory condition.

18. The use of an antisense peptide or polypetide as defined in any one of claims 1 to 13 in the manufacture of a medicament for use in the prophylaxis or treatment of a condition mediated by a cytokine.

19. The use as claimed in claim 18 wherein the cytokine is IL-1α or IL-1β, TNFα or IL-8.

20. The use as claimed in claim 19 wherein the condition is an inflammatory condition.

21. The use as claimed in claim 20 wherein the inflammatory condition is rheumatoid arthritis or septic shock.

22. The use as claimed in claim 19 wherein the cytokine is TNFα and the condition is cancer.

23. A pharmaceutical formulation comprising at least one antisense peptide or polypeptide, together with one or more pharmaceutically acceptable carriers, diluent or excipient.

24. A pharmaceutical formulation as claimed in claim 23, wherein the antisense peptide or polypeptide is as defined in any one of claims 1 to 13.

25. A method of treating a condition mediated by a cytokine which comprises administering to a patient an effective amount of an antisense peptide or polypeptide.

26. A method as claimed in claim 25 wherein the antisense peptide or polypeptide is as defined in any one of claims 2 to 13.

27. A method for the prophylaxis or treatment of an inflammatory condition which comprises administering to a patient an effective amount of an antisense peptide or polypeptide as defined in any one of claims 5 to 13.

28. A method for the treatment of cancer which comprises administering to a patient an effective amount of an antisense peptide or polypeptide as defined in claim 9 or claim 10.

		N																		C
IL-1 α	53	K	F	D	M	G	A	Y	*K*	*S*	*S*	*K*	*D*	*D*	*A*	K	I	..	T	V I
IL-1 β	41	V	F	S	M	S	F	V	*Q*	*G*	*E*	*E*	*S*	*N*	*D*	K	I	..	P	V A
IL-1 ra	45	K	I	..	..	D	V	V	..	..	..	..	..	..	..	P	I	E	P	H A

FIG. 1A

Q *G* *E* *E* *S* *N* *D*

5' CAA GGA GAA GAA AGT AAT GAC *3'* (IL-1β code)

3' GTT CCT CTT CTT TCA TTA CTG *5'* (Antisense code)

L *S* *F* *F* *T* *I* *V*

FIG. 1B

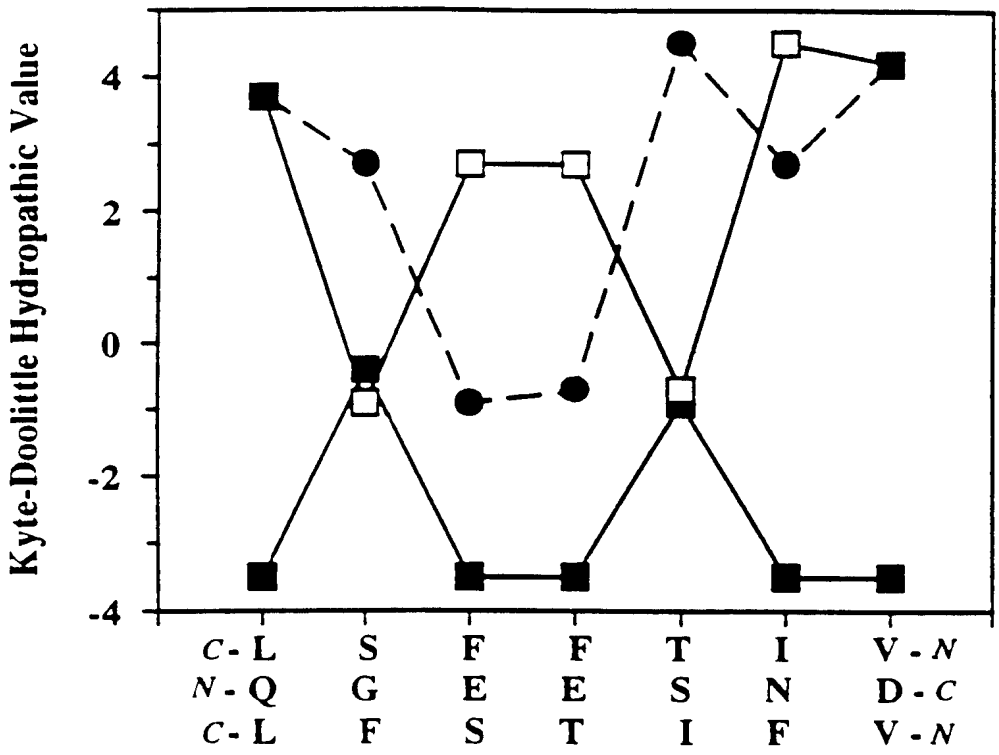

Kyte-Doolittle Hydropathic Value
4
2
0
-2
-4
C - L S F F T I V - N
N - Q G E E S N D - C
C - L F S T I F V - N

FIG. 1C

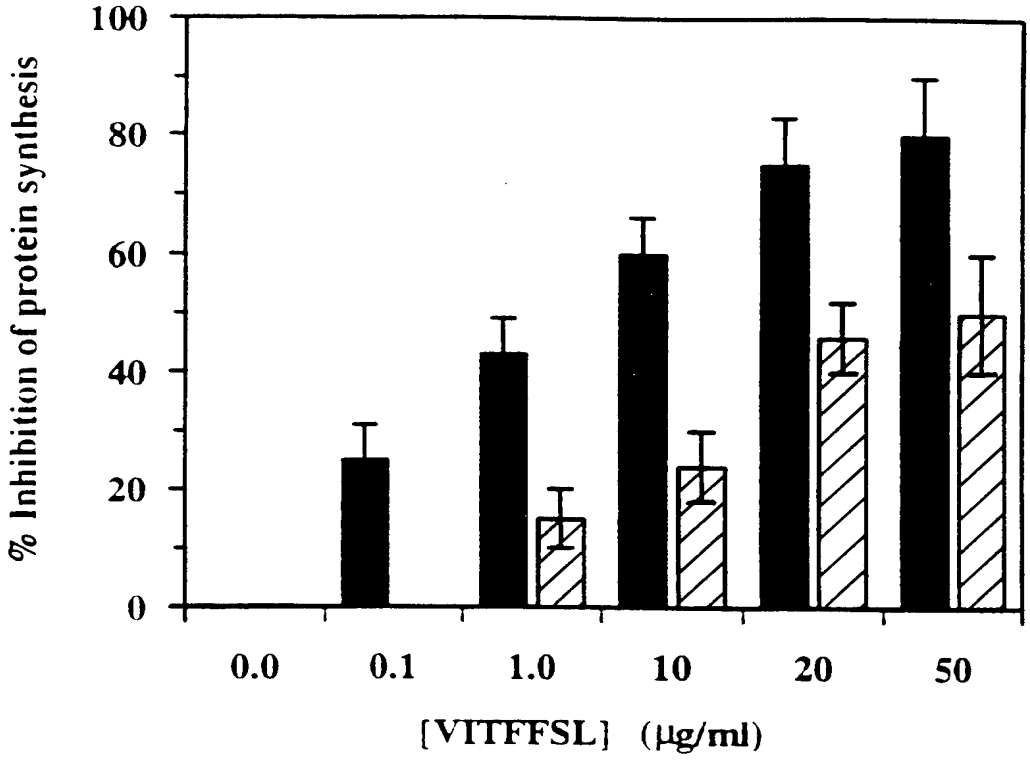

% Inhibition of protein synthesis
100
80
60
40
20
0
0.0
0.1
1.0
10
20
50
[VITFFSL] (µg/ml)

FIG. 2A

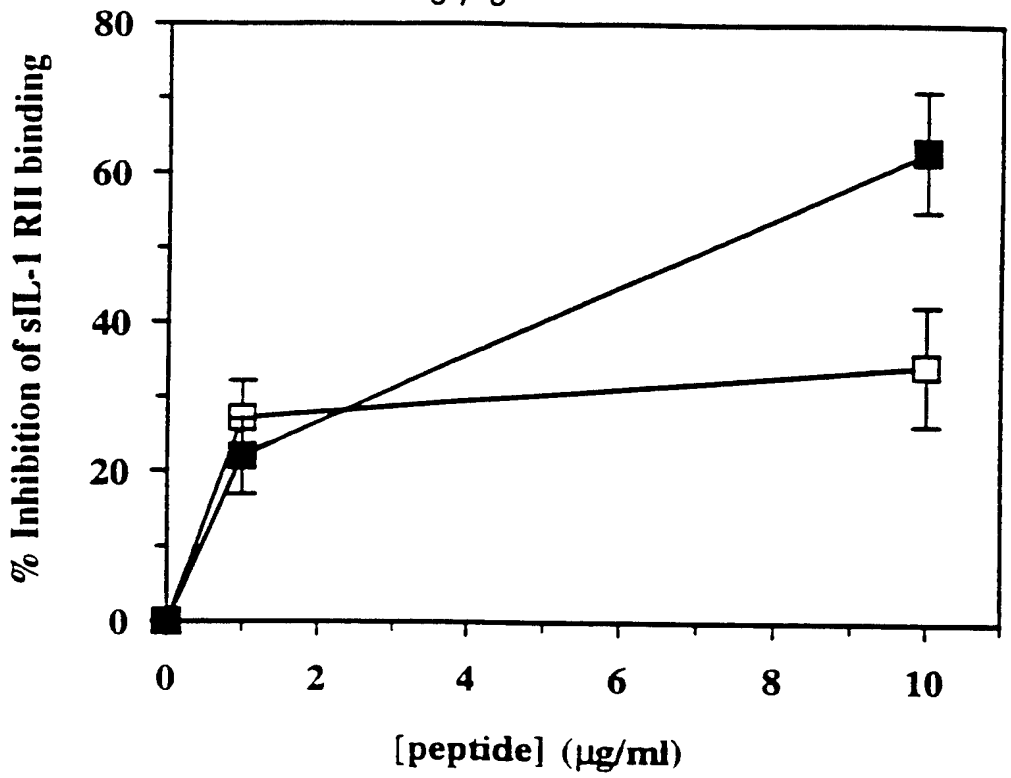

% Inhibition of sIL-1 RII binding
80
60
40
20
0
0
2
4
6
8
10
[peptide] (µg/ml)

FIG. 2B

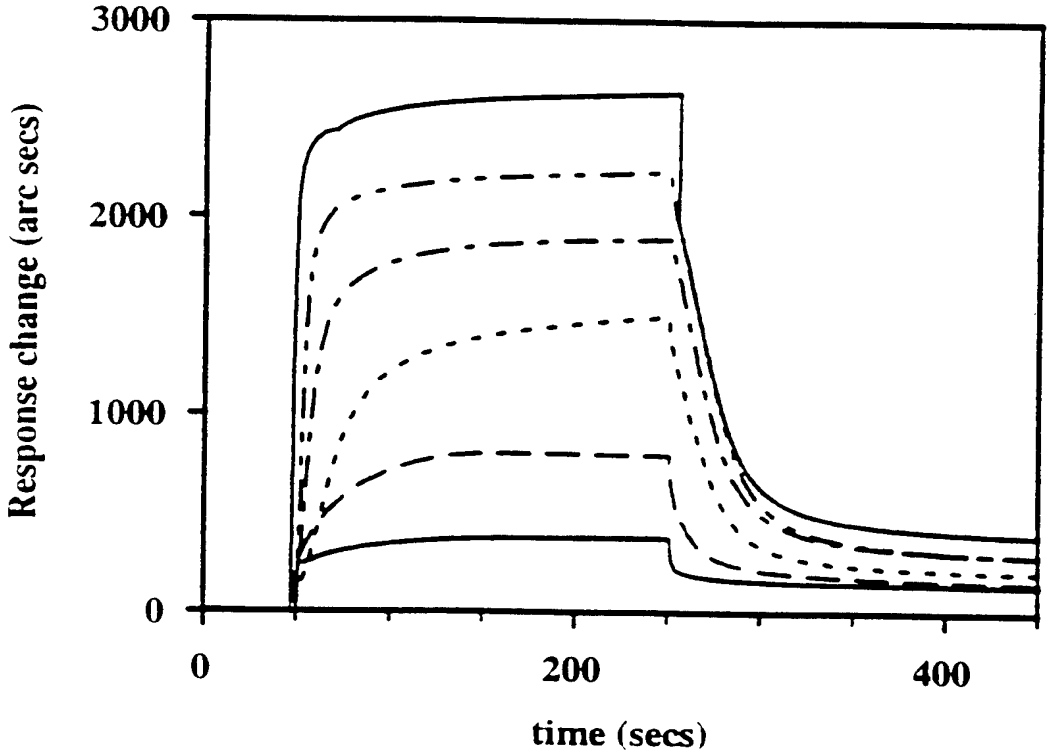

Response change (arc secs)
3000
2000
1000
0
0
200
400
time (secs)

FIG. 3A

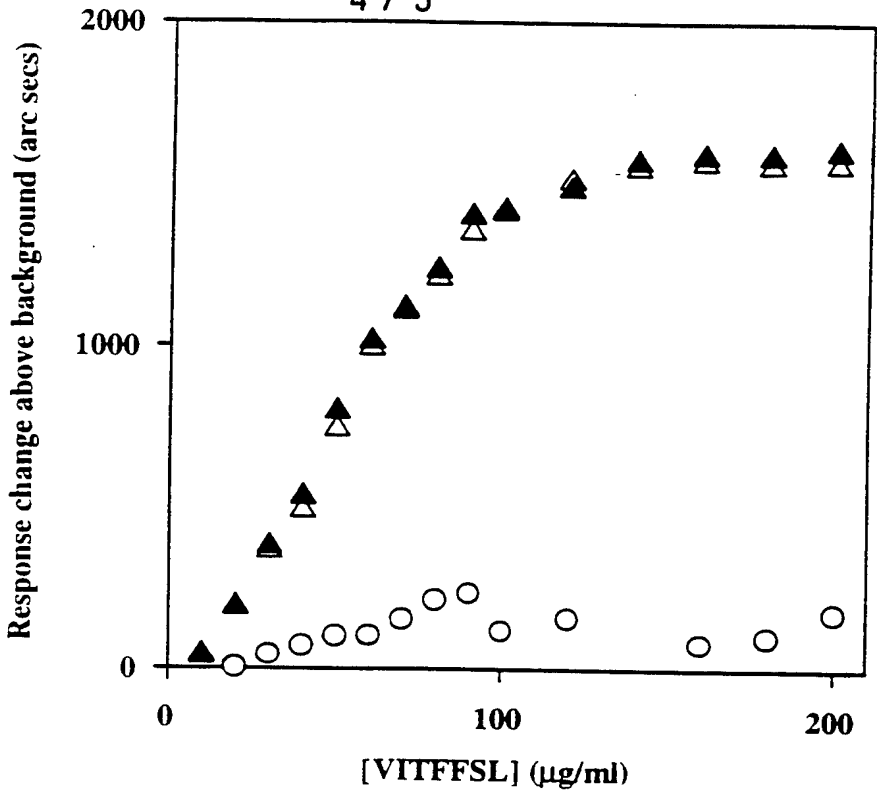
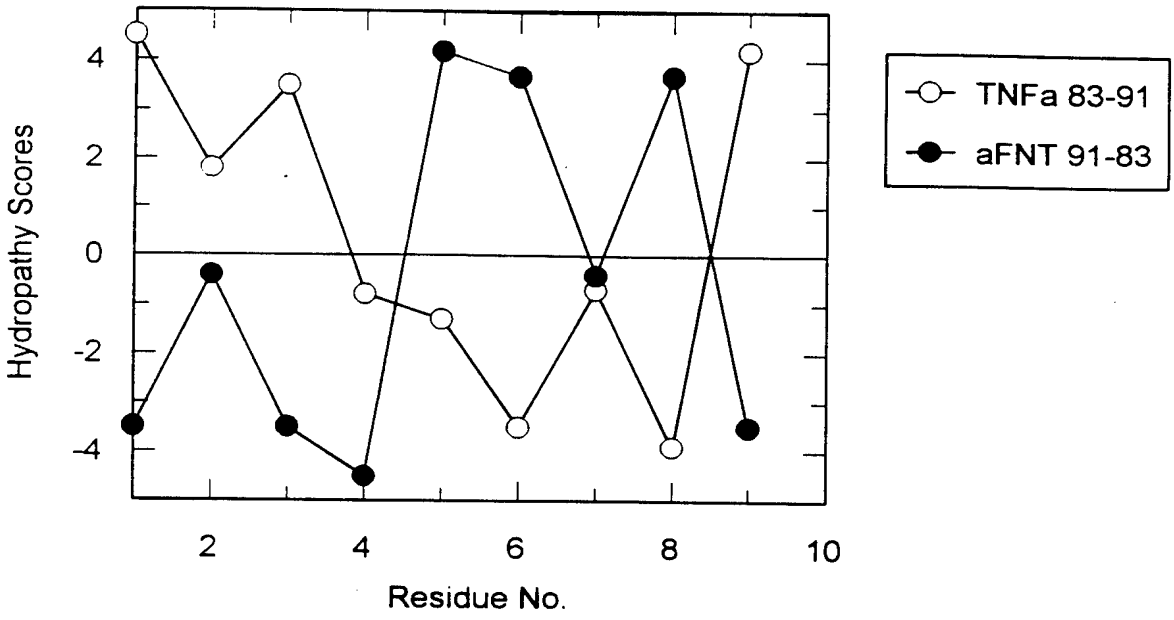
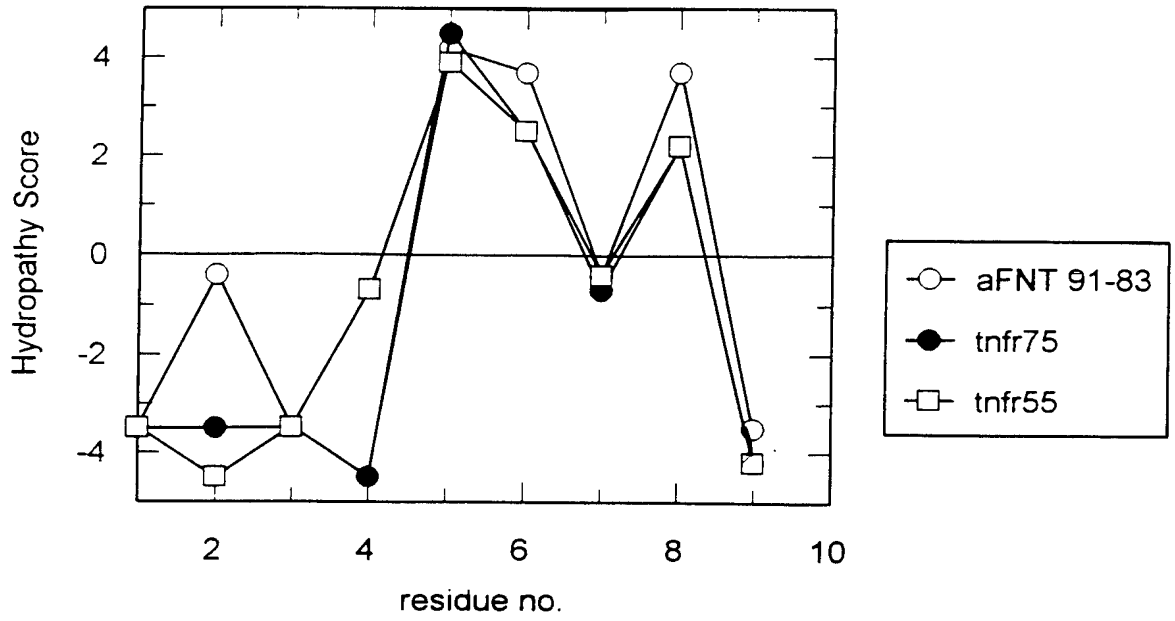


FIG. 3B



I A V S Y Q T K V TNFalpha
D L G L V R D G D aFNT

FIG. 4

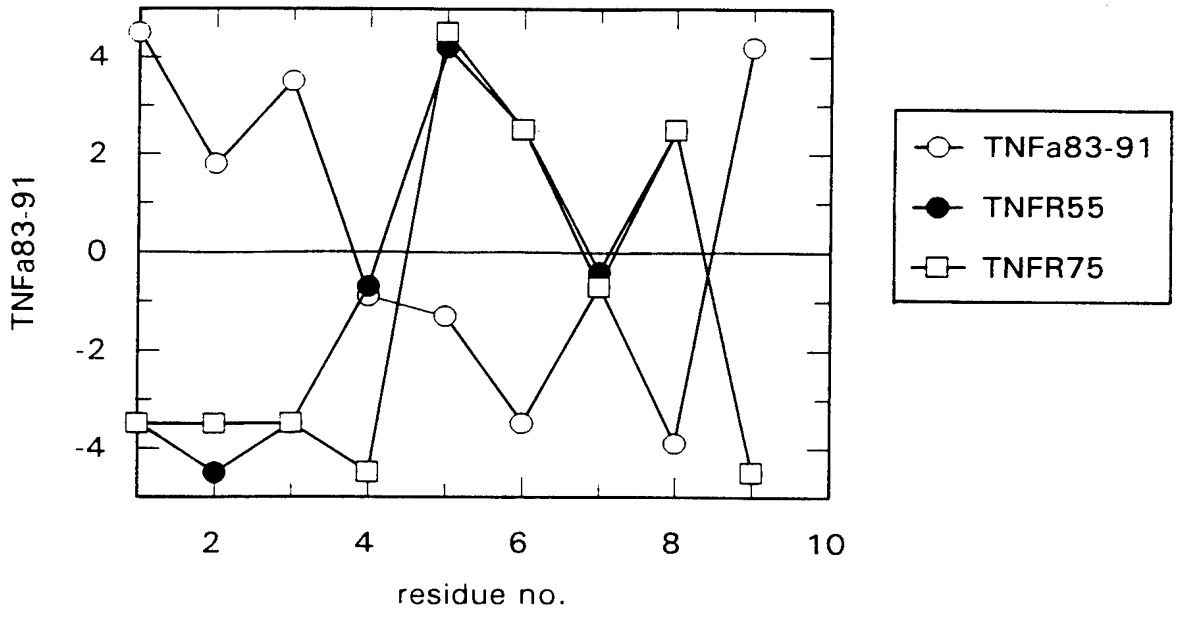


D G D R V L G L D aFNT

D R D T V C G C R TNFR p55 (91-99)

E Q N R I C T C R TNFR p75 (91-99)

FIG. 5



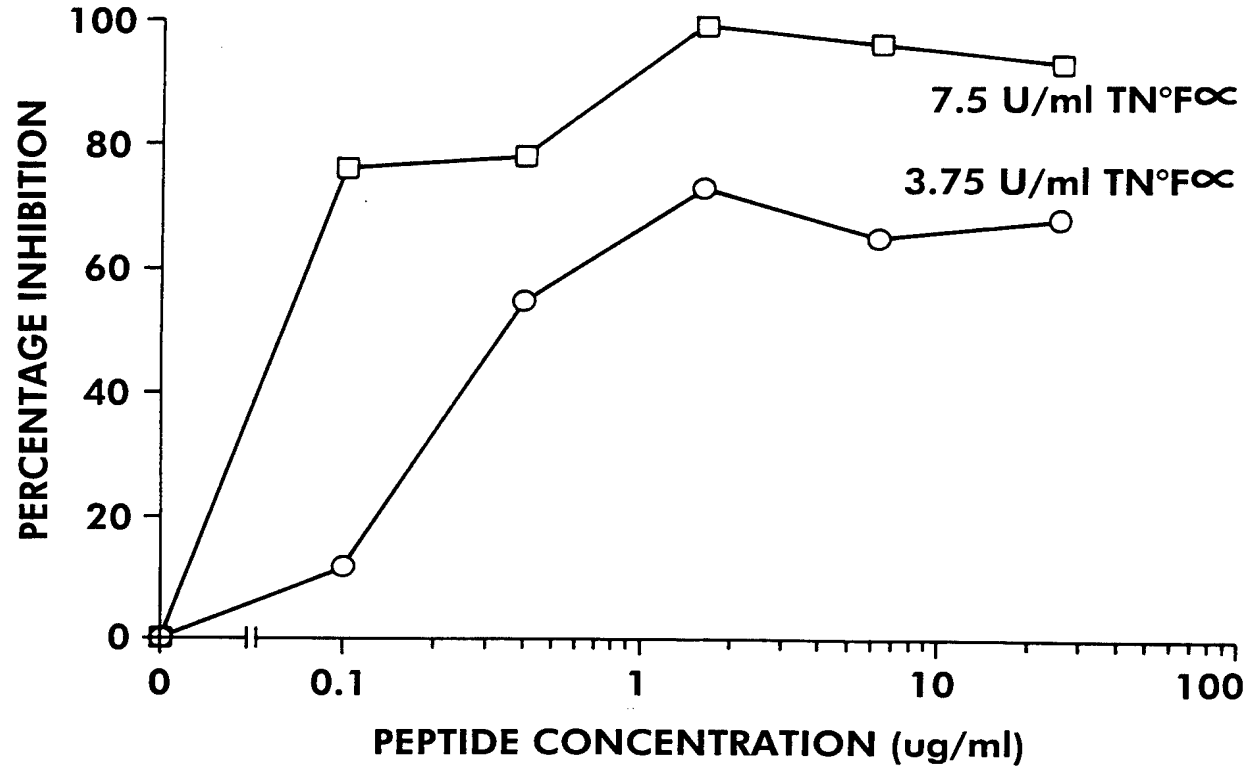
I A V S Y Q T K V TNFa 83-91

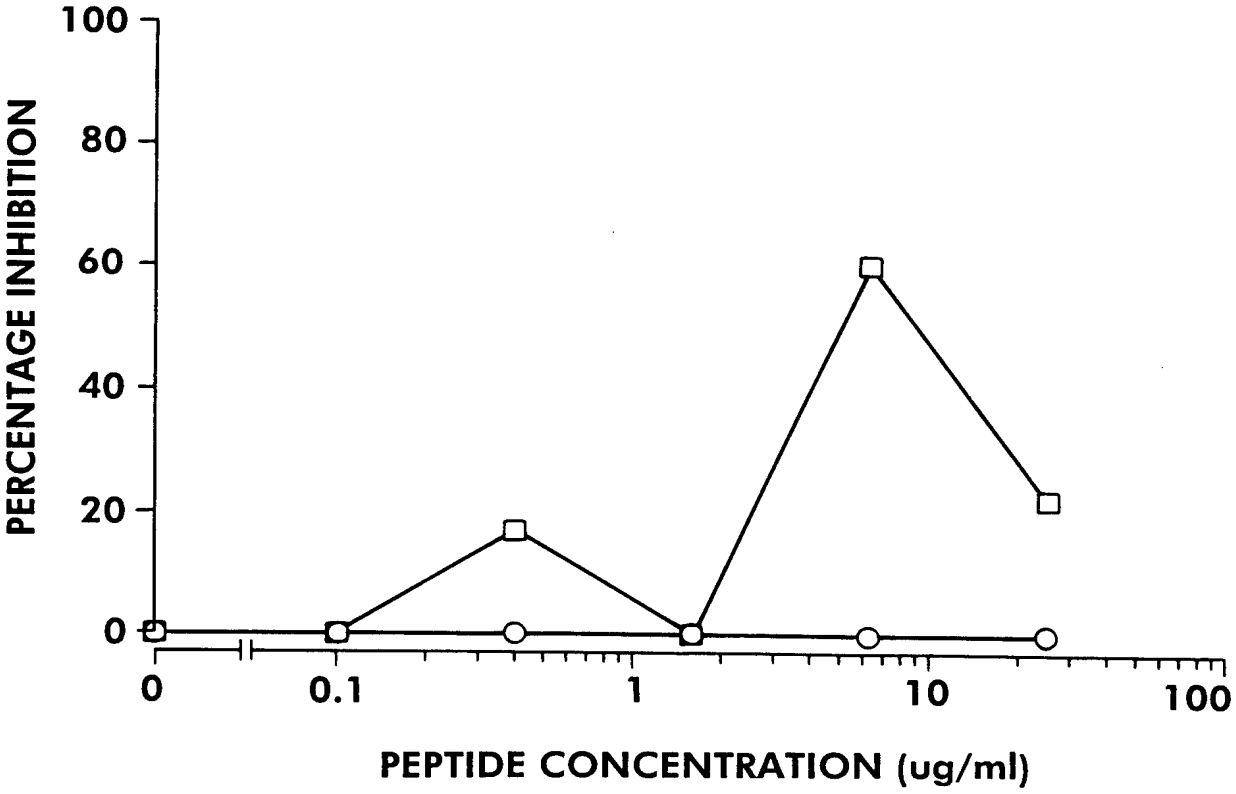
E Q N R I C T C R TNFR p55 91-99

D R D T V C G C R TNFRp75 91-99

FIG. 6

FIG. 7


100
80
60
40
20
0
PERCENTAGE INHIBITION
7.5 U/ml TNF∝
3.75 U/ml TNF∝
0
0.1
1
10
100
PEPTIDE CONCENTRATION (ug/ml)


100
80
60
40
20
0
PERCENTAGE INHIBITION
0
0.1
1
10
100
PEPTIDE CONCENTRATION (ug/ml)

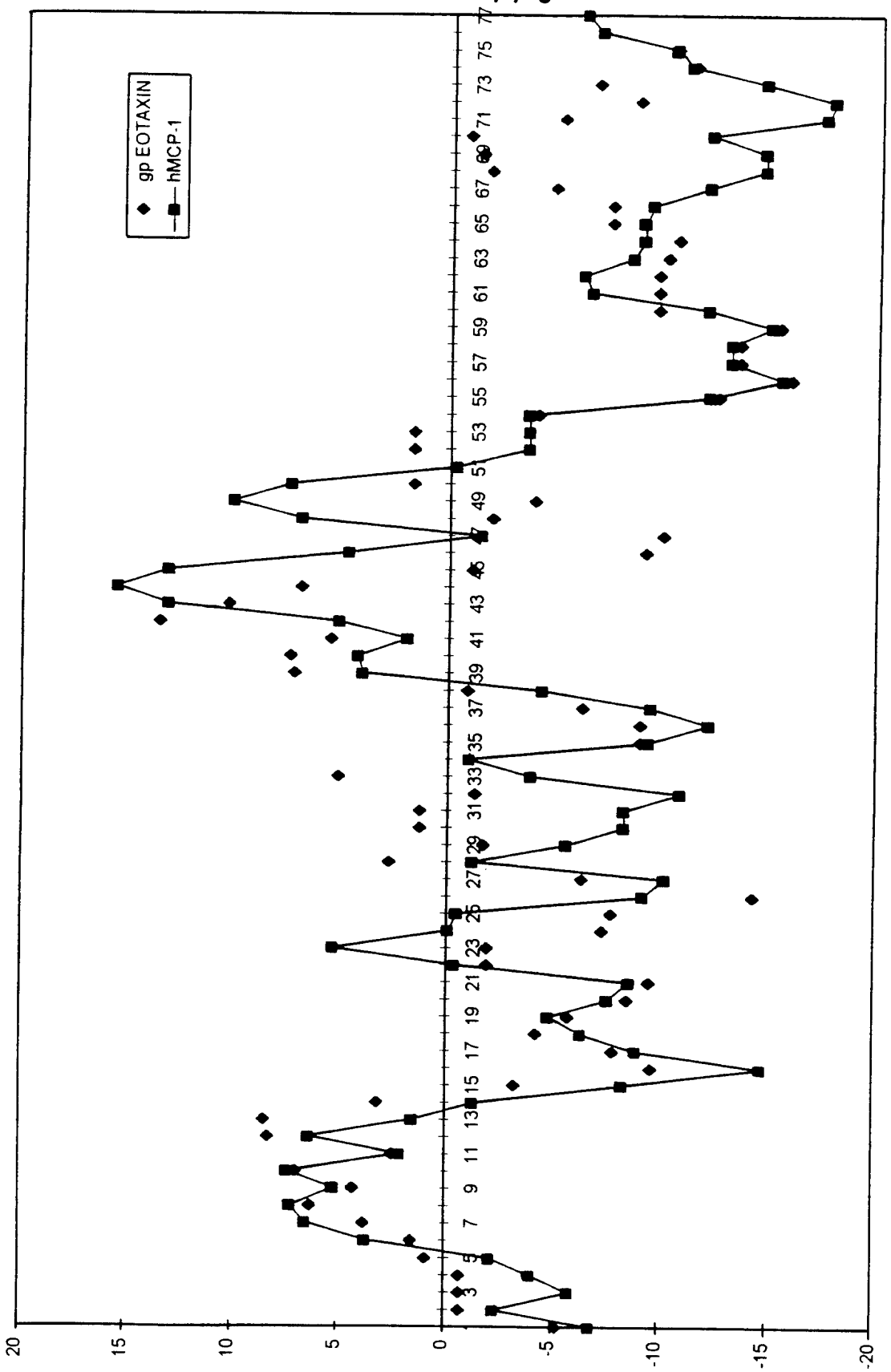


FIG. 8

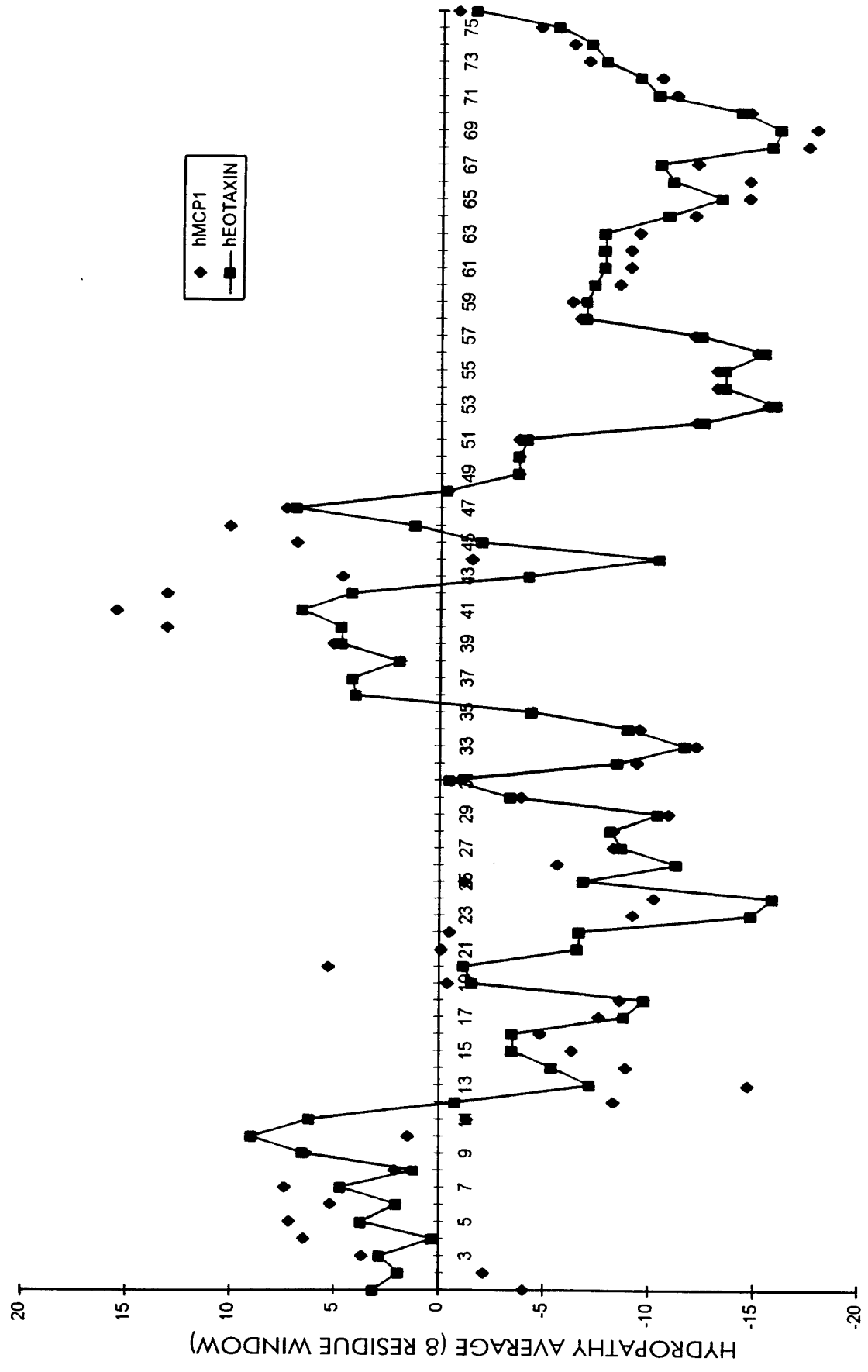
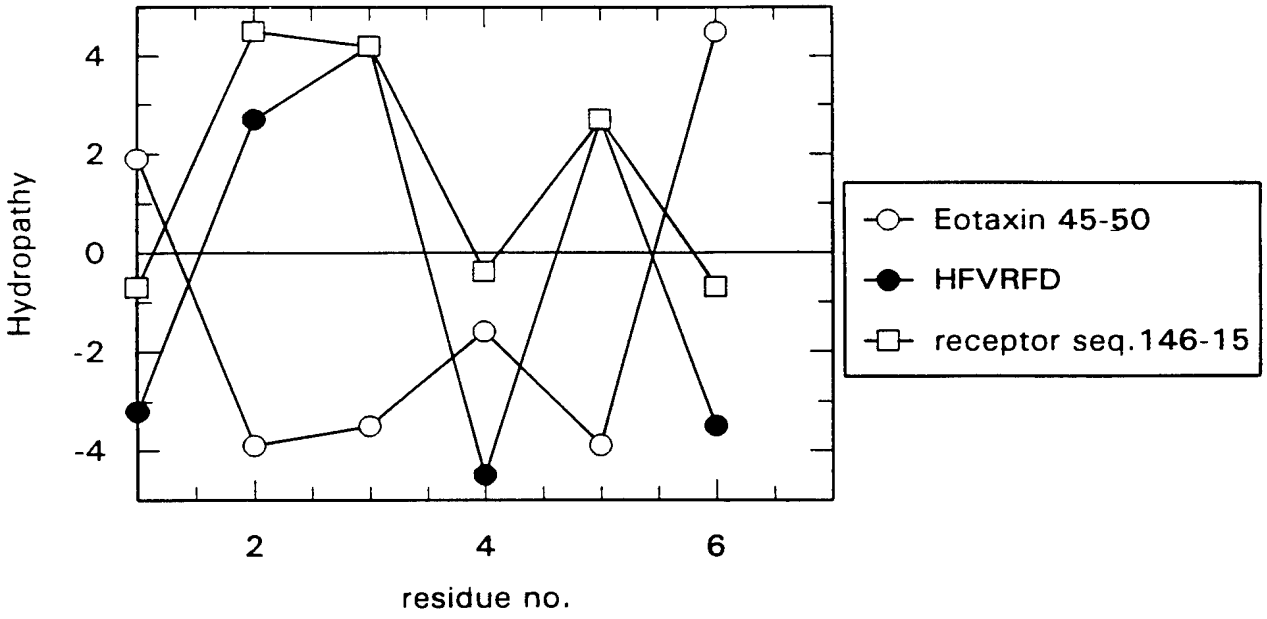

hMCP1
hEOTAXIN
HYDROPATHY AVERAGE (8 RESIDUE WINDOW)
20
15
10
5
0
-5
-10
-15
-20
3
5
7
9
13
15
17
21
23
27
29
33
37
39
41
47
49
51
53
55
57
59
61
63
65
67
69
71
73
75
RESIDUE NUMBER

FIG. 9



I K P D K M gp EOTAXIN 45 - 50
H F V R F D antisense peptide
T I V G F T CCCKR3 146 -152

FIG. 10

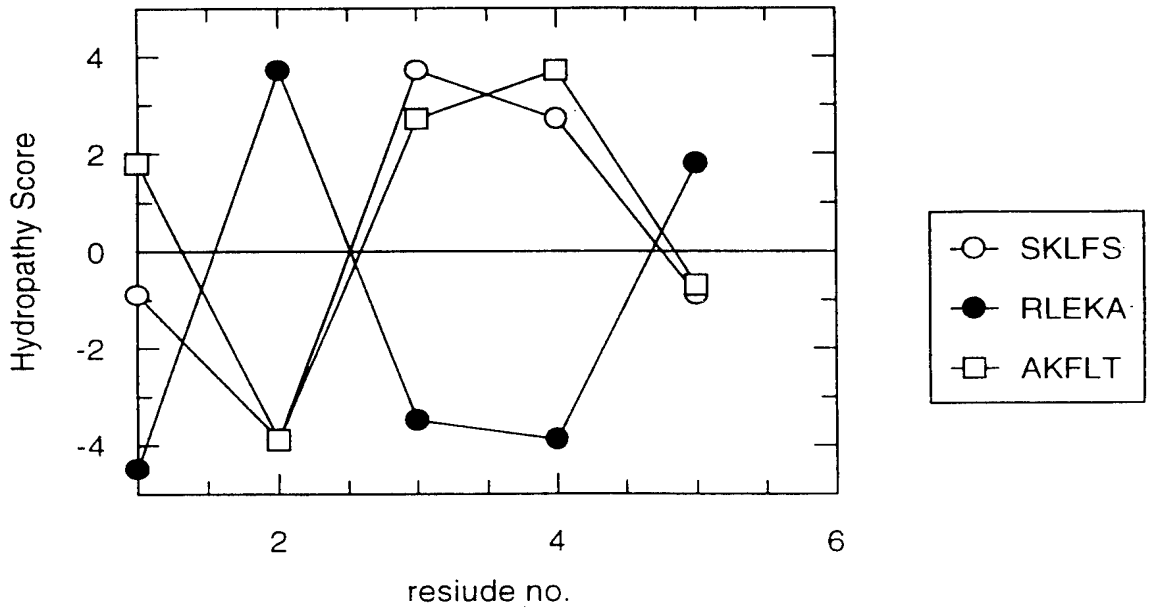


FIG. 11

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/62, C07K 19/00, 14/82, 14/52, 14/78, 14/47, 16/18, 16/28, 16/44, C12N 1/21, 1/19, 5/10, A61K 38/17, 39/395, G01N 33/50 // (C12N 1/21, 1:19)

A2

(11) International Publication Number: **WO 96/37621**

(43) International Publication Date: 28 November 1996 (28.11.96)

(21) International Application Number: PCT/EP96/02230

(22) International Filing Date: 23 May 1996 (23.05.96)

(30) Priority Data:
95107914.4 23 May 1995 (23.05.95) EP
(34) Countries for which the regional or international application was filed: DE et al.

(71) Applicant *(for all designated States except US)*: MORPHOSYS GESELLSCHAFT FÜR PROTEINOPTIMIERUNG MBH [DE/DE]; Frankfurter Ring 193a, D-80807 Munich (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PACK, Peter [DE/DE]; Franz-Wolter-Strasse 4/III, D-81925 Oberföhring (DE). HOESS, Adolf [DE/DE]; Wirtsbreite 21, D-83672 Warngau (DE).

(74) Agent: VOSSIUS AND PARTNER; P.O. Box 86 07 67, D-81634 München (DE).

(81) Designated States: CA, CN, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: MULTIMERIC PROTEINS

(57) Abstract

The present invention relates to the construction and use of small multimerization devices, preferably of human origin, which self-assemble fused functional domains to multimeric and multifunctional complexes. Multimerization devices of this invention do not significantly interfere with secretion, expression yields and the independent folding of functional domains, which are attached via flexible but protease-resistant linkers. Modular gene cassettes encoding functional domains, linkers and multimerization domain can easily be combined to a cistron encoding the multimeric protein. Translation in a suitable host results in self-assembly to multimers larger than dimers. In cases, in which one or both functional domains are not expressible in sufficient yields or native fold in the same expression host, multimeric proteins can be produced, in which one or both functional domains or the multimerization device are produced separately by, e.g., in vitro translation, peptide synthesis and/or refolding and subsequently, e.g., chemically coupled to the remaining part of the multimeric protein.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

MULTIMERIC PROTEINS

Field of the invention

The present invention relates to multimeric polypeptides which combine two or more functional domains in a structure which is capable of at least self-trimerization and which can be, for example, readily prepared in recombinant bacteria.

Background of the Invention

Multivalency is a prerequisite for a variety of macromolecular interactions such as binding of antibodies or lectins to specific targets, ligand recognition, activation or inhibition of receptors as well as cell adhesion.

Carbohydrate binding by lectins such as conglutinin or IgE-binding protein requires multivalency (Hsu et al., 1992, J. Biol. Chem. 267, 14167-14174; Andersen et al., 1992, J. Struct. Biol. 109, 201-27). Compared to the monovalent ligand, the trimeric hepatic lectin (Verrey & Drickamer, 1993, Biochem. J. 292, 149-155) binds the trimeric ligand with an 100 to 1000-fold higher affinity (Lee et al, 1992, Arch. Biochem. Biophys. 299, 129-136).

Multimerization of cell surface receptors as well as of their ligands often define specificity and affinity. The multivalent interactions of CD2 receptors of T-cells with their ligands LFA-3 on the cognate partner are necessary for cell-cell contact and T-cell activation by stimulatory cells (Moingeon et al., 1989, Immunol. Rev. 111, 111-144).

Multivalent binding of heparin to fibroblast growth factor receptors induces receptor dimerization and results in mitogenic signal transduction (Pantoliano et al., 1994, Biochemistry 33, 10229-10248). The concentration-dependent dimerization of the cytokine IL-2 results in high-affinity binding to the IL-2 receptor and formation of the high-affinity complex regulating the subsequent signal transduction (Lother et al., 1992, Growth factors 7, 117-129). Dimerization of CD4 induced by multivalent ligands is necessary to trigger the intracellular tyrosine kinase in T-cells (Langedijk et al., 1993, Thromb.-Res. 71, 47-60). The inhibition of IgE synthesis in allergic diseases requires the injection of at least bivalent anti-IgE immunoglobulins (Stämpfli et al., 1994, Eur. J. Immunol. 24, 2161-2167).

The enhancement of weak monovalent forces by multiplying the number of interactions is increasingly used for protein engineering and therapeutic applications. Synthetic pepides such as the bivalent hirudin-analogs "hirulogs" effectively inhibit clot-bound thrombin (Cannon et al., 1993, Am. J. Cardiol. 71, 778-782), a further optimization of related thrombin inhibitors being achieved by trivalency (Szewczuk et al., 1993, Biochemistry 32, 3396-3404). Multivalent glycopeptides inhibit binding of influenza viruses to their target cells (Unverzagt et al., 1994, Carbohydr. Res. 251, 285-301). The multivalent display of Tyr-Ile-Gly-Ser-Arg peptides strongly inhibits lung metastasis (Nomizu et al., 1993, Cancer-Res. 53, 3459-3461).

The synergistic gain in binding strength (functional affinity or avidity) through multivalent interactions is best characterized for binding of immunoglobulins to their specific antigens. The avidity effect is dependent on the affinity per binding site, flexibility and number of the

associated binding sites, number as well as distance of antigens within reach. As a rough estimate, the avidity is the product of affinities per monovalent binding event (Crothers & Metzger, 1972, Immunchemistry 9, 341).

Immunoglobulins such as monoclonal antibodies or heterologously expressed fragments belong to the most promising candidates for "targeted missiles" for screening and combat of tumours and any kind of pathogens in vivo (Colcher et al., 1990, J. Natl. Cancer Inst. 82, 1191-1197). The heterologous expression of so-called scFv antibody fragments, i. e. the variable domains of antibodies connected by a peptide linker (Bird et al., 1988, Science 242, 423; Huston et al., 1988, Proc. Natl. Acad. Sci. USA 85, 5879) combines several benefits.

The deletion of all constant immunoglobulin domains and the covalent linkage of the two variable domains (Fv) by a peptide linker (Huston et al., 1988, Proc. Natl. Acad. Sci. USA 85, 5879; Bird et al., 1988, Science 242, 423) results in a stable, single-chain immunoglobulin of minimal size with a single, complete binding pocket. The small size of the single-chain Fv fragment (scFv) provides improved expression yields (Skerra & Plückthun, 1991, Protein Eng. 4, 971), proteolytic stability, tissue penetration (Colcher et al., 1990, J. Natl. Cancer Inst. 82, 1191) and decreased nonspecific interactions or immunogenicity compared to complete antibodies. An alternative design links the two variable domains via engineered disulfide bridges at the VL-VH interface (dsFv; Glockshuber et al., 1990, Biochemistry 29, 1362).

The disadvantage of these recombinant fragments, however, is their short in vivo half-life and their inability to form

bivalent complexes such as IgG or higher oligomers such as IgM, in which the binding pockets are able to reach two or more distant antigens simultaneously. In medical applications such as tumour targeting, however, the need for efficient tissue penetration and tight binding means that ideally, an immunoglobulin-based structure should combine relatively small size with high affinity/avidity and sufficient half-live (Thomas et al., 1989, Cancer Res. 49, 3290 3296). To regain the avidity of whole antibodies and prolonged half-live in vivo, several techniques were recently developed to access small dimeric and thus bivalent immunoglobulins, although each suffers from substantial disadvantages.

If the interdomain linker between VH and VL in single-chain Fv fragments (Bird et al. , 1988, Science 242, 423-426) is sufficently short, formation of monovalent scFv fragments is disfavoured, resulting in intermolecular association of VH and VL domains of two or more fragments to form scFv-dimers or higher oligomers. In the so-called diabody (Holliger et al., 1993, Proc. Natl. Acad. Sci. USA 90, 6444), the rigid structure and close proximity of the binding sites does not allow flexibility or binding to distant antigens (Perisic et al., 1994, Structure 2, 1217). The formation of stable dimers or higher oligomers is highly dependent on the VH-VL interface of a particular fragment, so that the conversion of an scFv to a diabody is not possible for every given sequence, and the degree of oligomerization is not controllable (Whitlow et al., 1994, Prot. Eng. 7, 1017-1026).

The direct in vitro linkage of two fragments with C-terminal cysteines by oxidation (Carter et al., 1992, Bio/Technology 10, 163) or through chemically attached spacers (Kipriyanov

et al, 1994, Mol. Immunol. 31, 1047-1085; Cook & Wood, 1994, J. Immunol. Meth. 171, 227-237) requires laborious isolation and refolding of purified material with poor yields of functional immunoglobulins. Additionally, since mixed disulphide bond formation cannot be excluded, aggregation to unspecified multimers may occur. Non-quantitative disulphide formation further diminishes the overall yield.

The use of amphipathic helices such as bundle helices or leucine zippers C-terminally fused to scFv fragments for dimerization in E.coli (Pack & Plückthun, 1992, Biochemistry 31, 1579-1584) provided access to dimeric, bivalent "mini-antibodies" (PCT/EP93/00082). The antibody fragments assemble exclusively to functional dimers in vivo and can be purified in high yields in a single affinity chromatography step (Pack et al., 1993, Bio/Technology 11, 1271-1277). Based on artificial bundle helices or non-human zipper-sequences, this design of the C-terminally fused "association domains" is potentially hampered by the immunogenicity of such sequences when applied as human therapeutics. Furthermore, bundle helices used as association domains are not able to tetramerize fused antibody fragments, since the association forces of single bundle helices are too weak to stabilize a tetramer (Pack & Plückthun, 1992, Biochemistry 31, 1579-1584).

Modified and therefore "artificial" zipper sequences fused to antibody fragments are able to bring about tetramerization, but incompatibilities of the modified zipper sequence with folding and secretion lead to a considerably diminished yield of functional tetrameric material (Pack et al., 1995, J. Mol. Biol. 246, 28-34) and

the complex incorporates only one type of function or functional domain, namely one scFv specificity.

Previously, the CH2-CH3-CH4 domains of immunoglobulin M heavy chains have also been used as carriers of a single functional domain, namely cell surface receptor ligands (PCT Application WO 92/03569). The resulting fusion proteins were expressed in certain mammalian cells as multimers.

For several reasons, the use of CH2-CH3-CH4 domains of immunoglobulin M heavy chains as multimerization devices suffers disadvantages.

First, the expression of such fusion proteins is restricted to certain expression hosts and cellular compartments. The specific glycosylation of constant domains of immunoglobulins is considered to be important for assembly, structure (Rademacher & Dwek, 1984, Prog. Immun. 5, 95-112) function (Burton, 1985, Mol. Immunol. 22, 161-206) and proteolytic stability (Leatherbarrow & Dwek, 1983, FEBS Lett. 164, 227-239) of immunoglobulin constant domains (Matsuda et al., 1990, Mol Immunol. 27, 571-579; Dorai et al., 1991, Hybridoma 10, 211-217). The change or absence of glycosylation patterns, for instance when using a murine instead of a human expression host, results in structural changes and loss of complement activation (Lund et al., 1990, Mol. Immunol. 27, 1145-1153). The correct human glycosylation pattern on immunoglobulin constant domains is only achievable in certain mammalian cell cultures, namely of human origin. A different glycosylation pattern by expression in other than human cells would also increase the immunogenicity in medical applications.

Furthermore, the use of antibody-based association domains of the proposed type suffers from the serious disadvantage of its enormous size of 343 residues per monomer corresponding to three distinct immunoglobulin folding domains (CH2, CH3, CH4) per monomeric fusion protein, with 6 disulphide bridges to be formed (Dorai & Gillies, 1989, Nucleic Acids Res. 17, 6412-6417) or 60 in the decameric, IgM-like complex. Constructs of this type are very unlikely to be producible in bacteria, and there is no report of the functional expression of an immunoglobulin having three correctly folded constant domains in bacteria.

It is well known that the folding efficiency and expression yield of recombinant soluble material depends on the length of the peptide chain, the number of distinct folding domains per chain as well as the number of cis-prolines and correct disulphide bridges to be formed (Jaenicke, 1987, Prog. Biophys. Mol. Biol. 49, 117-237). In procaryotic expression hosts like E. coli, the decrease in yield of larger recombinant proteins is even more pronounced than in yeast or cell culture. As one aspect of this observation, for instance, the proteolysis of recombinant and "foreign" proteins, especially in inter-domain sequences, increases with the number of folding domains per chain. The formation of correct disulphide bridges is a rate-limiting step of folding (Darby et al., 1992, J. Mol. Biol. 224, 905-911) and a prerequisite for a stable immunoglobulin fold (Creighton, 1988, BioEssays 8, 57-53; Skerra & Plückthun, 1991, Prot. Eng. 4, 971-979). In reducing environments such as cytoplasmic compartments, the formation of a disulphide bridge could not be shown (Glockshuber et al., 1992, Biochemistry 31, 1270-1279).

Self-assembly of relatively small fusion proteins, preferably without the need of glycosylation or disulphide bridges in vivo, however, would overcome these general problems mentioned above and result in novel, well expressed complexes.

Bacteria such E. coli provide a fast generation turnover, cheap fermentation, well established molecular biology protocols such as fast access to site-directed mutations. Recently, the display of recombinant proteins on bacteriophages (Greenwood et al., 1991, J. Mol. Biol. 220, 821-827) proved to be a unique tool in combination with random mutagenesis for the generation of large libraries and selection for the most valuable protein mutants. Only small and well expressed multimerization devices, preferably peptidic multimerization devices, would allow the efficient construction of multimeric proteins comprising library-derived recombinant proteins as functional domains.

A further limitation of WO 92/03569 is that it does not teach the preparation of multimers having two or more distinct functional domains per fusion protein. Indeed, WO92/03569 only provides for one example of a certain receptor ligand to be assembled in the multimeric proteins when fused to the CH2-CH3-CH4 domains of immunoglobulin M heavy chain and expressed in certain mammalian cells.

None of the prior art can provide substances combining the features of a relatively small size, low immunogenicity, and high yields of functional material with the self-assembly of stable multimeric complexes larger than dimers having at least two distinct functions.

The additional fusion of a second functional domain C-terminal to a peptidic, self-assembling multimerization device via a second flexible linker significantly broadens the range of applications of multimeric proteins. Self-assembly of a bifunctional fusion protein with the help of a "central" multimerization device (domain1-linker1-device-linker2-domain2) would result in a completely novel recombinant complex having two functions in preferably multiple copies in one complex.

A wide range of combinations of two different functional domains is now within reach. The assembly of two or more different antibody fragments to multivalent complexes with valencies equal to/or more than 3 per specificity as well as combination of unrelated functional domains such as enzymes, toxins, cytokines, kinases, phosphatases, lectins, peptide hormones, cell adhesion proteins such as integrins, metal-binding domains, peptidic vaccines, bioactive peptides or soluble cell surface proteins such as the CD molecules of leucocytes or parts thereof in one multimeric complex are of considerable interest for diagnostic and therapeutic applications.

The drawbacks associated with the prior art are overcome by the present invention, which is concerned with the preparation of multifunctional proteins which show the following features:

(i) they can be prepared using standard recombinant micro-organisms, even though the molecular weight of the assembled protein exceeds that of proteins commonly expressed in bacteria,
(ii) low immunogenicity in humans

(iii) exists as multimers of valency equal to/or greater than three
(iv) provide proteins carrying two or more functional domains in a single multimeric structure.

The present invention surprisingly provides small peptidic multimerization devices which support two functional domains, which may be distinct from each other, as N- and C-terminal fusions, with all three subunits folding into functional forms, and which can be produced without the need for refolding, for instance, in recombinant bacteria.

Within a multimeric protein of this invention, which can be encoded and translated as a single, contiguous monomeric polypetide chain (domain1-linker1-multimerization device-linker2-domain2) with posttranslational self-assembly, such a peptidic multimerization device, however, must not greatly interfere with the folding and function of the N- and C-terminally fused domains, and it must self-associate in the steric context of the fused domains at both termini, even when the functional domains are of much larger size than the peptidic multimerization device. Peptidic multimerization devices of this invention provide for these properties.

It is well known in the art that not all kinds of polypeptides, i.e. functional domains can be produced in sufficient amounts and native fold in every expression host. This is due to differences in the folding and secretion environment, post-translational modifications such as glycosylation patterns and protease activities. However, the present invention also provides a solution to overcome these prior art problems. Namely, in these cases, a combined use of different expression hosts and/or in vitro synthesis has

to be used to create multimeric proteins with a central, peptidic multimerization device of this invention.

A certain functional domain, for instance, may require folding in the cytoplasm of a prokaryotic organism with subsequent refolding, but the second functional domain, with which the first functional domain should be combined does not properly fold in the cytoplasm of a prokaryotic organism but would need secretion in a mammalian host with subsequent glycosylation for a native fold.

Certain functional domains or parts therof can be produced in vitro in cell-free translation systems (Nishimura et al., 1995, J. Fermentation and Bioengineering 80, 403-405) or as completely synthetic polypeptides (Fitzgerald et al., 1995, J. American Chemical Society 117, 11075-11080). Polypeptides with more than 100 amino acid residues, for instance, are currently synthesized using Fmoc chemistry on solid phase (Roggero et al., 1995, Molecular Immunology 32, 1301-1309).

Using in vitro synthesis, either a part or a complete multimeric protein of this invention can be synthesized and subsequently coupled, e.g. chemically or enzymatically in vitro (e.g. Harris et al., 1973, FEBS Lett. 29, 189-192) or by oxidation of additional, free cysteins (Carter et al., 1992, Bio/Technology 10, 163) to the remaining parts, which are separately produced in a different expression system or expression host. For example, the multimerization core may be cheaply synthesized in E. coli whereas the linkers and functional domains may be produced in different expression systems such as CHO cells, or bioactive peptides as functional domains by in vitro synthesis.

The benefits of a peptidic multimerization device of this invention can be demonstrated using the example of scFv fragments as N- and/or C-terminally fused functional domains. The multimeric protein combines greatly improved properties such as multivalent binding and bispecificity with prolonged half-live in vivo making it therefore particularly suitable for a variety of therapeutic applications. Fullfilling the needs of small size for tissue penetration, monomeric scFv fragments do not show avidity effects and are rapidly excreted, mainly via the kidneys. Their half-live t(1/2 alpha) is around 10-times shorter than that of a whole, bivalent antibody and 2 to 3-times shorter than that of a dimeric scFv fragment (Haunschild et al., 1995, Antibody, Immunoconjugates and Radiopharmaceuticals 8, 111-127), diminishing the chance to find the specific antigen and bind thereto. Multimeric scFv fragments are expected to have an considerably higher half-live and circulation time in vivo than monomeric or dimeric scFv fragments. Additionally, they can bind multivalently and can combine two specificities in one multimeric complex (the N-terminal scFv fragment as a functional domain may have a different specificity than the C-terminal functional domain).

The present invention provides for multimeric proteins comprising scFv fragments as N-and C-terminal functional domains which, for instance, combine two distinct scFv fragments $(scFv1\text{-}linker1\text{-}device\text{-}linker2\text{-}scFv2)_{>2}$. This entirely new design of multivalent & bispecific immunoglobulins is of enourmous interest for diagnostic and therapeutic purposes, since it enables, for instance, multivalent targeting of tumor tissue (specificity scFv1) and the recruitment of cytotoxic T-cells (specificity scFv2) to the tumor (Perez et al., 1985, Science 316, 354-356).

Furthermore, the present invention generally provides for heterotetramers that combine a variety of functions in one molecule. For example, the heterotetramerization may be a consequence of the association of structurally different multimerization devices. Alternatively, they may be the result of differnt functional domains fused N- or C-terminally (via the same or different linkers) to said devices as is detailed hereinbelow. Thus, the peptidic multimerization domain of this invention can be based on parts of natural proteins with the ability to heteromultimerize, such as the the histone $(H3/H4)_2$ heterotetramer or the $(TAF_{II}42/TAF_{II}62)_2$ heterotetramer (Xie et al., 1996, Nature 380, 316-322), which both require less than 80 residues per subunit to heterotetramerize. Used as multimerization devices of this invention, 4 different functional domains (two domains N-terminally and two C-terminally fused via flexible linkers to the multimerization devices) can in principle self-assemble to a multimeric protein, in which 4 distinct functional domains appear twice.

It was surprisingly shown that the multimerization device fused to the linker and functional domain sequences as described in accordance with the present invention allows independent folding (Mack et al., Proc. Natl. Acad. Sci. USA 92, 7021-7025 (1995) and is essentially resistant to proteolytic attack which is not true for prior art constructs that employ a certain linker length. The linkers used in accordance with the present invention are preferably not larger than 15 amino acids.

The present invention also provides for a multimeric protein wherein the first functional domains are different for each

monomer and/or the second functional domains are different for each monomer. I this embodiment, the present invention also provides for the possibilities that the first and second functional domain(s) are the same or different.

Additionally, the present application provides for the possibility that the first and/or second functional domain(s) is/are fused to at least one further (functional) domain. Such constructs may advantageously comprise tags as functional domains at their N- or C-termini which may be used for purification purposes.

The present invention generally also provides for multimeric proteins of synthetic, semisynthetic or recombinant origin, wherein at appropriate locations cysteine residues are placed that are capable of forming cystine bridges within the multimerization device. Such cystine bridges confer a further enhanced stability on such multimeric proteins. The person skilled in the art is well aware of how such cysteines should be introduced into the multimerization devices and where they should be placed. For example, by appropriatly manipulating the DNA sequence of the invention, certain amino acid codons may be changed into cystein codons. Accordingly, thus modified DNA sequences are also comprised by the present invention.

The present invention also provides multimeric proteins and DNA sequences encoding the monomeric parts thereof, respectively, comprising appropriately designed additional cysteines as part of the multimerization device or the linker(s) to allow for intermolecular covalent linkage to additional functional domains, which are separately produced in vivo or in vitro and, for example, coupled to the

multimeric protein via oxidation of a free cystein to result in a covalent disulphide bridge.

The present invention also provides for multimerization devices, which are derived from libraries of randomized DNA sequences (EP 95 11 3021.0-2110 and EP-A 0 614 989). Such partially or fully randomized DNA sequences can be fused, for instance, to part of the gene III (gIII) of filamentous phage in an appropriate phagemid. After transformation of the phagemids into E. coli, gIII proteins can be produced as a fusion with a peptide encoded by one member of the library. After infection with helper phage, new phages are produced displaying the peptide of interest and incorporating the phagemid encoding for the displayed peptide. New multimerization devices can be identified by the interaction of the displayed peptide with a target peptide, which is, for instance, immobilized on a solid surface or fused to an infection mediating particle ("IMP").

Detailed description of the invention

The object of this invention is a family of multimeric proteins which combine two or more functional domains. The multifunctional proteins of the present invention are based on small preferably peptidic multimerization devices which at least trimerize, and which can be N- and C-terminally fused to functional domains to form novel multimeric and multifunctional polypeptide complexes. The invention also provides DNA sequences, vectors, encoding said multifunctional proteins as well as methods for their production in vitro and in vivo. The invention further relates to gene cassettes as well as to pharmaceutical and diagnostical compositions comprising the protein of the invention.

Specifically, the present invention provides DNA sequences which encode:

(i) a first functional domain,
(ii) a first linker sequence,
(iii) a multimerization device of preferably 30-80, and not more than 110 amino acids in length, which is capable of self-assembly to a trimer or higher order oligomer, and which is preferably of human composition,
(iv) a second linker sequence, and
(v) a second functional domain.

In this context, the term "functional domain" refers to an oligo- or polypeptide which has one or more functions, such as binding to a defined target substance, catalyzing reaction of a defined substrate, inhibiting the action of an enzyme, binding or blocking a receptor binding site or binding to a metal ion.

The items (i) to (v) mentioned above are preferably contiguous, and comprise a single open reading frame, allowing controllable expression of a single multi-functional protein from an appropriate vector. The illustration of the fusion principle of such a DNA sequence encoding a bifunctional, self-associating protein is shown in Fig. 1A.

The term 'linker sequence' is well known in the art. The linker used is not crucial for the present invention. Thus, in one construct, one and the same or different sequences can be used for the first and the second linker sequence.

A particular preferred embodiment of this invention is the use of a peptidic multimerization device based on a modified C-terminus of human p53 protein (Clore et al., 1994, Science 265, 386-391; Sakamoto et al., 1994, Proc. Natl. Acad. Sci. USA 91, 8974-8978; Soussi et al., 1990, Oncogene 5, 945-952).

The gene product of the human p53 tumor suppressor gene (Hollstein et al., 1991, Science 253, 49-53) contains distinct modular sequences for trans-activation (Unger et al., 1992, EMBO J. 11, 1383-1386), DNA-binding (Erlandson & Verdine, 1994, Chemistry & Biology 1, 79-84) and tetramerization (Clore et al., 1994, Science 265, 386-391; Sakamoto et al., Proc. Natl. Acad. Sci. USA 91, 8974-8978). The sequence reponsible for tetramerization of the p53 (residues 319-360; Fig.2) can be separately expressed in E. coli (Studier et al., 1990, Methods Enzymol. 185, 60-64) resulting in a stable, symmetric 20 kD tetramer, containing helical as well as beta-strand secondary structures (Jeffrey et al., 1995, Science 267, 1498-1502).

The particular preferred p53-based multimerization device surprisingly allows self-assembly to a tetrameric complex with two different functional domains fused to its termini (Examples 1 - 3), where each of the fused functional domains can be more than eight times larger than the device (Example 2). Suprisingly and contrary to what the person skilled in the art would have expected, the human p53-based peptidic multimerization device has, in several cases, not greatly or significantly interfered with secretion, independent folding and function of the domains fused via flexible linkers, since the properties of both functional domains can be detected in the tetrameric complex directly isolated from the cell extract (Example 1). The small size and human origin of the p53 based peptidic multimerization device enables a significantly reduced immunogenicity in human compared to artificial or non-human self-associating sequences.

In a further preferred embodiment, the engineered peptidic multimerization device comprises self-associating, modified human platelet factor 4 (PF4; Zhang et al., 1994, Biochemistry 33, 8361-8366). PF4 (Fig. 4) is an abundant human serum protein of 101 residues secreted by platelets. It can be heterologously expressed in E. coli and self-assembles into tetramers with antiparallel beta-sheet like structures (Zhang et al., 1994, Biochemistry 33, 8361-8366).

The extracellular human protein TSP4, forming the basis of a further preferred embodiment of this invention, is a member of the thrombospondin family, which modulates the attachment of a variety of cells (Lawler et al., 1993, J. Cell. Biol. 120, 1059-1067). Residues 209 to 273 of human TSP4 (Fig. 21) show a high degree of homology to residues 20 to 83 of the

cartilage oligomeric protein (CMP or COMP). This homologous region is responsible for pentamerization of COMP and TSP4 (Efimov et al., 1994, FEBS letters 341, 54-58; Jenkins et al., 1990, J. Biol. Chem. 265, 19624-19631). Pentamerization of functional domains can be achieved by C- or N-terminal fusion to a TSP4-based peptidic multimerization device according to Examples 1 and 2.

In a preferred embodiment of the invention, the multimerization device is derived from a randomized DNA library in a suitable host, from which DNA members are identified by their ability to encode multimerizing peptides. Means and methods for such identification are well known to the person skilled in the art; see, e.g., EP 95 11 3021.0, or EP-A 0 614 989, which are incorporated herein by reference.

The present invention also provides for proteins, preferably fusion proteins, multimeric proteins consisting e.g. of such fusion proteins as well as methods of producing multimeric proteins of this invention. Such a method relates to the production of a protein comprising

a) a first functional domain,
b) a first linker sequence,
c) a preferably peptidic multimerization device of not more than 110 and preferably 30-80 amino acid residues in length, which is capable of self-assembly to a trimer or higher order oligomer, and which is of predominantly human composition,
d) a second linker sequence, and
e) a second functional domain,

wherein at least one of a) to e) is

i) produced recombinantly, and, if all of a) to e) are produced recombinantly, at least two of the constituents are produced by different host cells; and/or
ii) produced synthetically; and/or
iii) produced semisynthetically; and/or
iv) produced by in-vitro translation; and

wherein at least two of a) to e) are combined by enzymatic and/or chemical coupling thereby giving rise to the complete protein, and preferably to the complete multimeric protein.

For therapeutic purposes, it is often desirable that proteinaceous substances display the minimum possible immunogenicity. Accordingly, the present invention provides for multimeric polypeptides as described above in which at least one of said amino acid sequences, functional domains, or further (poly)peptides is of human origin. For example, the extracellular origin of the peptidic multimerization device based on parts of human PF4 or COMP is anticipated to provide for low immunogenicity in humans.

In a further preferred embodiment of this invention, the peptidic multimerization device comprises parts of the TATA Box binding protein associated factors (TAFIIs), preferably of human origin. Drosophila dTAFII42 and dTAFII62 are able to form hetero-tetrameric complexes resembling the (H3/H4)2 hetero-tetrameric core of the chicken histone octamer (Xie et al., 1996, Nature 380, 316-322). The human homologues to dTAFII42 and dTAFII62 are hTAFII31 (Hisatake et al., 1995, Proc. Natl. Acad. Sci. USA 92, 8195-8199) and hTAFII80 (Histake et al., 1995, Proc. Natl. Acad. Sci. USA 92, 8195-8199; Weinzierl et al., 1993, EMBO J. 12, 5303-5309). Residues 13 to 87 of the human hTAFII31 and residues 10 to 82 of the human hTAFII80 can be used as a heterotetramerization device of this invention. Different

nomenclature may exist for identical hTAFII-proteins, such as hTAFII80 is also known as hTAFII70 (Hisatake et al., 1995, Proc. Natl. Acad. Sci. USA 92, 8195-8199).

In a further preferred embodiment of this invention, the peptidic multimerization device comprises parts of the histone H3 and H4 proteins, preferably residues 67 to 134 of human histone 3 (H3) and residues 29 to 95 of the human histone 4 (H4) (Fig 12a, b). The human H3 and H4 are highly homologous to the chicken H3 and H4, which resemble the structurally similar dTAFII42/dTAFII62 hetero-tetramer (Xie et al., 1996, Nature 380, 316-322).

In order to provide sufficient distance and flexibility between the peptidic multimerization device and fused functional domains, the peptidic multimerization device is linked via preferably hydrophilic, preferably flexible but protease-resistant polypeptide linkers to the functional domains. In a preferred embodiment of this invention, one or both of the linkers comprise protease-stable inter-domain linkers of human proteins (Argos, 1990, J. Mol. Biol. 211, 943-958), which provide sufficient flexibility and hydrophilicity. In a further preferred embodiment, the linker sequence is rich in the amino acid residues serine, threonine, glycine and proline, which provide an extended structure, rotational freedom, hydrophilicity as well as stability against proteases (Argos, 1990, J. Mol. Biol. 211, 943-958). In a particular preferred embodiment of the invention, the linker is a human antibody hinge sequence, particularly the TPLGTTTHT sequence derived from the upper hinge region of the human IgG3 antibody (Fig.4). In this context, the term antibody hinge sequence refers to an oligopeptide which links CH1 and CH2 domains and is responsible for flexibility, rotational freedom and long

range of the Fab arm to bind simultaneously to distant antigens (Dangl et al., 1988, EMBO J. 7, 1989-1994)

The present invention provides for many different types of functional domains to be multimerized. Thus, they may

a) bind to a defined target substance, or
b) catalyze reaction of a defined substrate, or
c) inhibit the action of an enzyme, or
d) bind or block a receptor binding side, or
e) bind to a metal ion.

Preferred are also cases in which the sequences of at least one of the C- or N-terminal functional domains fused to the multimerization device comprise at least a part of a member of the immunoglobulin superfamily. In a particularly preferred embodiment, said functional domain is an antibody scFv fragment. This aspect of the invention provides multivalent antibody-like molecules, which can be of utility when simultaneous multivalent interaction with antigen is a prerequisite of recognition and stable binding to the specific antigen. This is the case, for instance, with carbohydrate antigens on a cell surface, against which individual antibody binding sites show poor affinity.

Also preferred are cases in which one or more functional domains fused to the peptidic multimerization device possess biological activity other than that of a member of the immunoglobulin superfamily. By way of example, the present invention provides for the targeted multimerization of enzymes, toxins, cytokines, kinases, phosphatases, lectins, peptide hormones, cell adhesion proteins such as integrins, metal-binding domains, peptidic vaccines, bioactive peptides or soluble cell surface proteins such as the CD molecules of leucocytes or parts thereof in one multimeric complex.

A preferred embodiment of this invention relates to constructs where the functional domains are bioactive peptides. Certain peptides derived from amphipathic loop structures of LPS-binding proteins (Hoess et al., 1993, EMBO J. 12, 3351-3356) are able to neutralize endotoxin. LPS often occurs in multimeric forms and it is therefore desirable to have access to multivalent forms of short (10-15 residues) LPS-binding peptides to ensure effective binding and neutralization of endotoxin. The present invention provides a method to express and multimerize several short peptides (Fig. 16) fused to the multimerization device. The peptides can be fused either to the N- or to the C-terminus (Fig. 17, 18) of the assembly domain via the peptide linkers. A fusion to both termini, e.g., to a p53 based multimerization device, results in a self-assembling octavalent complex of bioactive peptides. Alternatively, a combination with other functional domains can be achieved by their fusion to the opposite terminus of the peptidic multimerization device, resulting in a self-assembling complex displaying four bioactive peptides and four further functional domains.

The invention enables complex multimeric and multifunctional polypeptides to be constructed. By way of example, the second functional domain fused to the peptidic multimerization device can be taken from the list of cytokines, toxins, enzymes, kinases, phosphatases, lectins, peptide hormones, cell adhesion proteins such as integrins, metal-binding domains, purification devices, in particular peptides which are able to bind to an independent binding entity, peptidic vaccines, bioactive peptides, preferably of 5 to 15 amino acid residues, soluble cell surface proteins such as the CD molecules of leucocytes, DNA binding domains, transcription factors and growth factors or parts thereof.

A preferred embodiment of this invention is the use of decorsin as a functional domain. Decorsin is a 39 residue protein of the leech Macrobdella decora (Fig.13). It acts as a potent antagonist of the platelet glycoprotein IIb-IIIa (Seymour et al., 1990, J.Biol.Chem. 265, 10143-10147). According to example 2, the decorsin can be fused C- or N-terminally to the peptidic multimerization device (Fig. 3, 11, 12, 20, 22) and expressed in a suitable host. In arterial thrombotic diseases, the translated self-assembling multivalent decorsin complex may act as a powerful antithrombotic agent. Optionally, the N-terminal fusion of an anti-fibrin antibody fragment to a peptidic multimerization device, which carries a C-terminally fused decorsin, may result in targeting of the multimeric complex to blood clots.

The present invention provides for a standarized system of gene cassettes encoding functional domains, linkers and peptidic multimerization devices. Modular gene cassettes can easily be combined via unique restriction enzymes to a cistron (Fig. 1A). The cistron encodes for a self-associating protein, in which each monomeric subunit consists of an optional signal sequence for secretion, a first functional domain, a first linker sequence, an assembly domain of 30-110 amino acids in length, a second linker sequence and a second functional domain.

In a preferred embodiment said gene cassettes are directly cloned as a cistron into standard expression vectors enabling replication, under a regulatable promoter, translation and optionally secretion of the cistron-encoded protein. Host cells transformed with at least one vector or

vector cassette of the invention can be used for the preparation of said multimeric polypeptides.

In a further preferred embodiment said host cell is a mammalian, preferably human, yeast plant or bacterial preferably an E.coli cell.

The invention further provides for methods for the production of said multifunctional proteins, which methods comprise culturing a host cell of the invention in a suitable medium, and recovering said multimeric polypeptide produced by said host cell.

The invention further provides for methods for the separate production of parts of the multimeric protein such as the peptidic multimerization domain or/and one or both functional domains, and multimeric proteins and monomers thereof produced by said method which methods comprise at least one of peptide synthesis, in vitro translation, expression in hosts other than the hosts in which the remaining part of the multimeric protein is produced (e.g. by species, cell ionic or strain difference), and refolding with chemical coupling, e.g. subsequent or enzymatic, to the separately produced remaining part of the functional protein to result in a polypeptide that assembles to a multimeric protein of this invention.

The invention further provides for the introduction of appropriately placed additional cysteins to allow covalent intermolecular linkage of the multimerization device during or after self-assembly.

The invention further provides for the introduction of appropriately placed additional cysteins to allow linkage of

further functional domains, for example, by oxidation to intermolecular disulfide bridges.

In further preferred embodiments, the present invention provides for pharmaceutical and diagnostic compositions comprising the multimeric polypeptides described above, said pharmaceutical compositions optionally comprising a pharmaceutically acceptable carrier.

Finally, the invention provides for a kit comprising one or more vector cassettes useful in the preparation of said multimeric polypeptides.

The invention is now illustrated by reference to the following examples, which are provided for the purposes of illustration only and are not intended to limit the scope of the invention.

EXAMPLES

Example 1: Formation of tetrameric complex incorporating four copies anti-Le^Y scFv antibody fragments and four copies of a metal-binding domain by fusion of the two different functional domains to the 5'- and 3'-termini of a small tetramerization device based on human p53 via flexible linkers.

The peptidic multimerization device of this example is based on a modified C-terminus of human p53 (Soussi et al., 1990, Oncogene 5, 945-952; Clore et al., 1994, Science 265, 386-391; Fig. 2). A gene cassette (Fig. 3) synthesized by recursive PCR (Prodromou & Pearl, 1992, Protein Eng. 5, 827-829) encodes residues 319-360 of human p53 flanked by a SG-peptide incorporating an in-frame unique 5' restriction site (MroI) and a short C-terminal GGSGGAP linker incorporating an in-frame AscI restriction site at the 3'- terminus of the gene cassette encoding the peptidic multimerization device. The synthetic modular MroI-AscI gene cassette is 5' ligated to an EcoRI-MroI cassette encoding the human IgG3 upper hinge (Dangl et al., 1988, EMBO J. 7, 1989-1994; Fig. 4). The ligation product, a EcoRI-HindIII gene cassette encoding the tetramerization device, can be used for 5' (N-terminal) as well as 3' (C-terminal) fusions to genes encoding recombinant proteins or functional domains such as cytokines, toxins, immunoglobulins, enzymes, kinases, phosphatases, lectins, peptide hormones, cell adhesion proteins such as integrins, metal-binding domains, purification devices, in particular peptides which are able to bind to an independent binding entity, peptidic vaccines, bioactive peptides, preferably of 5 to 15 amino acid residues, soluble cell surface proteins such as the CD

molecules of leucocytes, DNA binding domains, transcription factors and growth factors (see further examples).

A cistron of the fusion protein of this example is constructed by the ligation of the following modular gene cassettes (5' -> 3') to a cistron und under a lac promoter/operator system in an appropriate E. coli expression plasmid (Fig. 6a, Ge et al., 1995, Antibody Engineering 2nd edition, C.A.K. Borrebaeck, ed. Oxford University Press, pp 229-266, Glockshuber et al., 1990, Biochemistry 29, 1362-1367; Skerra et al., 1991, Bio/Technology 9, 273-278):

- a XbaI-EcoRV cassette encoding an ompA signal sequence
- an EcoRV-EcoRI cassette encoding a LeY carbohydrate-binding scFv antibody fragment
- an EcoRI-MroI AscI gene cassette encoding a flexible upper hinge of human IgG3 (Fig. 4)
- a MroI-AscI encoding a peptidic multimerization device derived from the human IgG3 resuidues human p53 and a flexible GGSGGAP linker incorporating the in-frame AscI-site (Fig. 3)
- a AscI-HindIII gene cassette encoding a functional domain for purification purposes which complexes metal ions.

The EcoRV-EcoRI cassette encoding the LeY carbohydrate-binding scFv antibody fragment MSL-5 was constructed by PCR amplification of the V-region cDNA of a hybridoma (Bradbury et al., 1995, Antibody Engineering 2nd edition, C.A.K. Borrebaeck, ed. Oxford University Press, pp 295-361) secreting anti-LeY antibodies. Appropriate PCR primers were used to introduce the EcoRV and EcoRI restriction sites and to assemble the scFv fragment with a 15 residue $(Gly_4Ser)_3$ interdomain linker in a VL to VH polarity (Ge et al., 1995,

Antibody Engineering 2nd edition, C.A.K. Borrebaeck, ed. Oxford University Press, pp 229-266).

After transformation of the expression plasmid into an E. coli JM83 strain and induction of the lac promotor with IPTG at room temperature in shaking cultures, the translated fusion protein (MSL5-P53-His) is secreted into the periplasm and self-assembles into a tetrameric protein of this invention with four scFv fragments and four metal binding domains in one complex ("multibody"). In addition to applications such as incorporation of radioactive metal ions into a tetravalent anti-tumour complex, the metal binding function of the C-terminal domain can also be used for purification purposes. For example, the complex is isolated by immobilized metal ion chromatography IMAC (Lindner et al., 1992, Methods: a companion to Methods in Enzymology 4, 41-56; Fig. 5). The metal binding property of the tetrameric complex is demonstrated by the success of the purification procedure.

The peptidic multimerization device is compatible with secretion and folding of the anti Le^Y-binding scFv fragment, so that no reduction in functional yields (compared to the monovalent scFv) is seen. Unexpectedly, the tetrameric and thus tetravalent fusion protein showed an improved expression behaviour compared to the monomeric scFv in this example, probably due to a reduced toxicity.

The tetramerization of the bifunctional fusion protein to a stable complex is shown by size exlusion at 0.5 micromolar concentration (Fig. 6B).

The characterization of binding kinetics to a Le^Y-coated biosensor surface was performed by surface plasmon resonance on a BIAcore (Pharmacia; Pack et al., 1995, J. Mol. Biol. 246, 28-34). All measurements were carried out in HBS buffer (Pharmacia) with a flow rate of 5 µl/min at 25°C using one CM5 sensor cell of a biosensor machine (BIAcore, Pharmacia). LeY-BSA was covalently immobilized on the dextran matrix of a CM5 sensor chip using the standard amine immobilization procedure. Following activation of the chip surface with 50 mM N-hydroxysuccinimide and 200 mM N-(dimethylaminopropyl)-N'-ethylcarbodiimide, a LeY-BSA stock solution (155 µg/ml) in PBS was diluted with an equal volume of sodium acetate (1 M, pH 3.0) and injected twice to give a high density surface (7000 RU) of immobilized LeY-BSA. For analysis of the MSL5 proteins, 20 µl of each sample was injected at various concentrations (IgM: 0.3 - 14 nM, scFv: 0.37 - 3.75 µM; dimeric mini-antibody: 0.12 - 2.4 µM and multimeric MSL5-p53-His: 0.08 - 1.07 µM). Dissociation was induced by injection of HBS buffer. After each measurement, the chip surface was regenerated with 10 µl of 2 M guanidinium HCl diluted with an equal volume of HBS buffer.

The overlay plot of dissociation curves (Fig. 6c) reveals the second function of the tetrameric complex, namely the binding to a specific target, and illustrates the gain of avidity (functional affinity) by tetramerization of the scFv fragment antibody fragment. Compared to a monovalent scFv fragment and the dimeric construct scdHLX (PCT/EP93/00082), the tetrameric and thus tetravalent protein of this invention (MSL5-p53-His or "multibody") shows a significant enhancement of binding affinity to a surface covered with Le^Y-antigens. The gain in avidity can clearly be seen by the prolongation of the dissociation rate from milliseconds (scFv) to several minutes (MSL5-P53-His, Fig 6C).

The modular system of gene cassettes encoding the different association domains allows the facile generation of multivalent and bifunctional proteins with defined valency and steric properties. The multimeric protein of example 1 with four immunoglobulin binding sites can be an ideal design for applications where fast excretion and poor functional affinity of a monovalent fragment does not result in a significant, long-lasting enrichment at the target.

Furthermore, the dependence of the avidity (or "functional affinity") on the actual antigen density (Pack et al., J. Molecular Biology 246, 28-34) will allow a distinction between different tissues having different antigen densities. LeY is found in high densities on colon and breast cancer cells, but in low densities on a variety of normal cells (Sakamoto et al., 1986, Cancer Res. 46, 1553 - 1559). Since antibodies or immunotoxins with higher intrinsic affinity against LeY and related antigens also bind to cell surfaces with low antigen density, significant binding to normal tissues is observed (Trail et al., 1993, Science 261, 212 - 215). In principle, proteins of this invention with high avidity (by the multiplicity of binding events per molecule on a sufficient antigen densitiy) but comparatively low intrinsic affinities (affinity per monovalent interaction, occuring also on low densities) are unable to bind to cells with low densities. Therefore, they will give rise to more selective targeting of cells with over-expressed antigen, which favor multivalent binding and will not cause unwanted side-effects by binding to normal cells with low antigen density.

Example 2: Formation of tetrameric complex incorporating four copies anti-phosphocholine scFv antibody fragments and

four copies of a a human IL2 by fusion of the two different functional domains to the termini of a tetramerization device based on human p53 via flexible linkers.

According to example 1, a multimeric fusion protein is encoded by a cistron, which comprise modular gene cassettes encoding a signal sequence, a scFv antibody fragment as a N-terminal functional domain, a first flexible linker based on human IgG3 hinge, a tetramerization domain based on the human p53, a second linker and human IL2 as a C-terminal functional domain (Fig.8,9).

After translation and secretion in a suitable host according to example 1, the multimeric protein H11-p53-huIL2 is purified by the ability of its N-terminal functional domain to bind specifically to the antigen phoshocholine on a affinity column (Glockshuber et al., 1990, Biochemistry 29, 1362-1367). After this purification step, the function of its C-terminal domain, human IL2, was tested in a proliferation assay of cytotoxic T-lympocytes (CTLL-A).

3 x 10^6 cultivated CTLL-A cells, which can only grow in the presence of functional IL2, are extensively washed with IL2-free phosphate buffer (20 mM NaH_2PO_4, 100 mM NaCl, pH 7.4.) 2.5 % FCS, centrifuged and resuspended in IL2-free standard media with 10 % FCS. Aliquots of aproximately 1.5 x 10^4 cells in a 50 microliter volume are pipetted into each well of a 96-well microtiter plate. Increasing amounts of recombinant human IL2 for calibration (0.5 - 40 U; Sigma) as well as a dilution series of the phosphocholine-affinity purified H11-p53-huIL2 protein (10 ng - 150 ng) were co-incubated in duplicates for each concentration. After 24h at 37°C, the activity of active mitochondrial dehydrogenase (reflecting the number of living cells) is determined with MTT according

to Mossmann (1983, J. Immunological Methods 65, 55-57). The IL2 activity of the C-terminal functional domain of the multimeric protein (Fig. 9) is calculated using the calibration curve.

The proliferation of CTLL-A cells in the presence of the multimeric protein, which was purified with the help of the anti-phosphocholine scFv as the N-terminal functional domain, clearly demonstrates the activity of the huIL2 as the C-terminal functional domain.
The surprising result, that the central multimerization domain allows independent native folding of the distinct domains with subsequent self-assembly, illustrates the potential of this invention.

The central (-linker1-device-linker2-) part of the multimeric protein serves, apart from its multimerization function, as a very long and protease-stable "spacer", allowing independent folding of the considerably (up to 6 times) larger fused domains. A conventional, extremely long linker of the same length as the central (-linker1-device-linker2-) part but without its secondary structure is probably much more protease-sensitive, as it is typical for unstructured peptides, which are well accessible for the catalytic pocket of proteases (Argos, 1990, J. Mol. Biol. 211, 943-958).

Example 3: Formation of tetrameric and bifunctional complex incorporating four copies of anti-ESL-1 scFv antibody fragments and four copies of a detection tag which in addition is used as a metal-binding domain by fusion of the two different functional domains to the termini of a small tetramerization device based on human p53 via flexible linkers.

The multimerization device of this example is described in Example 1.

A cistron of this fusion protein is constructed by the ligation of the following modular gene cassettes (5' -> 3') to a cistron under a lac promoter/operator system in an E. coli expression plasmid according to example 1:

- a XbaI-EcoRV cassette encoding an ompA signal sequence
- an EcoRV-EcoRI cassette encoding ESL-1 binding scFv antibody fragments
- an EcoRI-MroI AscI gene cassette encoding a flexible upper hinge of human IgG3 (Fig. 4)
- a MroI-AscI encoding a multimerization device derived from the human IgG3 residues human p53 and a flexible GGSGGAP linker incorporating the in-frame AscI-site (Fig. 3)
- a AscI-HindIII gene cassette encoding a functional domain for purification purposes which complexes metal ions and detection purposes of being specifically recognized by an antibody (Fig. 5).

The ESL-1 binding scFv antibodies were derived from a human combinatorial antibody library using phage display. After three rounds of panning 12 clones different in the sequence of the CDR3 heavy chain were selected and cloned via XbaI and EcoRI in to an appropriate expression vector (Fig. 6a). The human combinatorial antibody library construction and the panning of the ESL-1 binding antibodies is described in EP application 9511 3021.0-2110 "Protein libraries", the teachings of which are incorporated herein by reference. After transformation of the expression plasmids into an E. coli JM83 strain and induction of the lac promoter with IPTG at 30°C in shaking cultures for 4 h, the translated fusion

proteins were secreted into the periplasm as self-associating tetramers (scFv-linker1-device-linker2-purification domain) without the need of refolding or chemical cross-linking. For lysis, the cells (4 ml) were collected by centrifugation, re-dissolved in 800 µl PBS pH 7.4 and sonicated for 90 sec. After another centrifugation step the supernatant was used for Western Blot analysis and binding experiments (ELISA). In parallel, conventional monomeric fusion proteins consisting of scFvs directly fused to the purification domain were prepared accordingly.

For characterizing the amounts of soluble secreted protein a Western Blot analysis was performed. 20 µl of lysed bacterial supernatant was loaded on a 12.5 % polyacryl amide gel. After blotting the gel, the nitrocellulose filter was blocked with 3 % milk powder. In addition to using the metal binding function of the C-terminal domain for purification purposes the C-terminal domain was used as a detection tag specifically recognized by an anti-His antibody (Dianova).

It is demonstrated that, similar to the case with the LeY binding antibody (Example 1), the multimerization device is compatible with secretion and folding of the anti-ESL-1 binding scFv fragments, so that no significant reduction in functional yields is seen in this example as judged from the amounts of soluble protein detected in the Western Blot.

The characterization of binding activity to surface bound ESL-1 reveals the second function of the tetrameric complex, namely the binding to a specific target. For example, in an ELISA experiment, the antigen (ESL-1, 5µg/ml) was coated overnight at room temperature to a microtitre plate. After blocking with 3 % milk powder, the antigen was incubated for 2 h at room temperature with 200 µl cell lysate either

expressing monomeric or tetrameric ESL-1 binding antibody fragments fused to the His-detection tag. After washing with PBS pH 7.4 (5x) the detection of ESL-1 binding was assayed by incubation with a first anti-His antibody (Dianova) and a second antibody linked to alkaline phosphatase and directed against the first antibody. Each measurement was performed in duplicate (Fig. 10).

The binding experiments of the ESL-1 binding antibody fragments reveal both functions of the tetrameric complex, namely the binding to a specific target and the detection of bound complexes, and illustrates the gain of avidity by tetramerization of the N-terminal functional domain, namely the scFv fragment antibody fragment. Compared to a monovalent scFv fragment the tetravalent and bifunctional proteins of this invention (each incorporating four copies of anti-ESL-1 scFv fragments) show a significant enhancement of measurable binding activity to a surface covered with ESL-1-antigens up to 10 or more times (9D5, 9D11, 9D31). In some cases, only binding of the multimeric and bifunctional complex of this invention is detectable (9D7, 9D9, 9D10) due to the synergistic effect of multiple binding sites (avidity-effect).

Example 4: Hetero-tetramerization device based on parts of the human histones H3 and H4

A synthetic gene containing the genetic information for the part of human histones H3 (Fig. 12a) and H4 (Fig. 12b) responsible for hetero-tetramerization are constructed with suitable restriction sites for 5' and 3' ligations according to Fig. 3, 20 and 22. The H3 and H4 based devices can be used for the hetero-tetramerization of multimeric proteins incorporating up to four different functional domains such

as cytokines, toxins, immunoglobulins, enzymes, kinases, phosphatases, lectins, peptide hormones, cell adhesion proteins such as integrins, metal-binding domains, purification devices, in particular peptides which are able to bind to an independent binding entity, peptidic vaccines, bioactive peptides, preferably of 5 to 15 amino acid residues, soluble cell surface proteins such as the CD molecules of leucocytes, DNA binding domains, transcription factors and growth factors. The hetero-tetramerization requires either co-expression of the two fusion proteins (domain1-linker1-device1 (e.g. H3) -linker2-device2 and domain3-linker3-device2 (e.g. H4) -linker4-device4) with the use of a suitable dicistronic expression system (Ge et al., 1995, Antibody Engineering 2nd edition, C.A.K. Borrebaeck, ed. Oxford University Press, pp 229-266) or separate expression with subsequent mixture and in vitro assembly. Instead of human histones H3 and H4, the hetero-tetramerization device can be based on sequences derived from the human $hTAF_{II}31$ and $hTAF_{II}80$ (Fig. 11 a,b).

Example 5: As a functional domain, the platelet aggregation inhibitor decorsin (Fig. 13) is either encoded by a N-terminal EcoRV-EcoRI (Fig. 14) or as C-terminal AscI-HindIII gene cassette (Fig. 15) and fused to a peptidic multimerization device of this invention via flexible linkers according to examples 1, 2 and 3.

Example 6: Formation of octavalent anti-LPS peptides (Fig. 16) by fusion to a tetramerization device based on the human p53.

The bioactive peptides can be fused as a gene cassette with suitable restriction sites either to the N-terminus (Fig.

17) or to the C-terminus (Fig. 18) of the assembly domain via the peptide linkers according to examples 1, 2 and 3.

Example 7: Tetramerization device based on the human serum protein PF4.

A synthetic gene containing the genetic information for PF4 (Fig. 19) and suitable restriction sites for 5' and 3' ligations (Fig. 20) can be used for the tetramerization of multimeric proteins incorporating functional domains such as cytokines, toxins, immunoglobulins, enzymes, kinases, phosphatases, lectins, peptide hormones, cell adhesion proteins such as integrins, metal-binding domains, purification devices, in particular peptides which are able to bind to an independent binding entity, peptidic vaccines, bioactive peptides, preferably of 5 to 15 amino acid residues, soluble cell surface proteins such as the CD molecules of leucocytes, DNA binding domains, transcription factors and growth factors according to Examples 1 to 3.

Example 8: Pentamerization device based on human TSP4 (Fig. 21), which can be used for the pentamerization of multimeric proteins incorporating functional domains such as cytokines, toxins, immunoglobulins, enzymes, kinases, phosphatases, lectins, peptide hormones, cell adhesion proteins such as integrins, metal-binding domains, purification devices, in particular peptides which are able to bind to an independent binding entity, peptidic vaccines, bioactive peptides, preferably of 5 to 15 amino acid residues, soluble cell surface proteins such as the CD molecules of leucocytes, DNA binding domains, transcription factors and growth factors according to Examples 1, 2 and 3.

CLAIMS

1. A DNA sequence encoding:
   a) a first functional domain,
   b) a first linker sequence,
   c) a multimerization device of not more than 110 and preferably 30-80 amino acids in length, which is capable of self-assembly to a trimer or higher order oligomer, and which is of mammalian and predominantly human composition,
   d) a second linker sequence, and
   e) a second functional domain.

2. The DNA sequence according to claim 1 in which said multimerization device comprises at least a part of the human p53 protein.

3. The DNA sequence according to claim 1 or 2 in which said peptidic multimerization device comprises amino acids 319 to 360 of the human protein p53.

4. The DNA sequence according to claim 1 in which said peptidic multimerization device comprises at least a part of the human PF4 protein.

5. The DNA sequence according to claim 1 in which said peptidic multimerization device comprises at least a part of the human TSP-4 protein.

6. The DNA sequence according to claim 1 in which said peptidic multimerization device comprises at least a part of the human COMP protein.

7. The DNA sequence according to claim 1 in which said peptidic multimerization device comprises at least a part of human thrombospondin.

8. The DNA sequence according to claim 1 in which said peptidic multimerization device comprises at least part of $dTAF_{II}42$ or $hTAF_{II}31$.

9. The DNA sequence according to claim 1 in which said peptidic multimerization device comprises at least part of $dTAF_{II}62$ or $hTAF_{II}80$.

10. The DNA sequence according to claim 1 in which said peptidic multimerization device comprises at least part of histone 3, preferably human histone 3.

11. The DNA sequence according to claim 1 in which said peptidic multimerization device comprises at least part of histone 4, preferably human histone 4.

12. The DNA sequence of any one of claims 1 to 11 encoding appropriately designed additional cysteines as part of the multimerization device to allow for intermolecular covalent linkage of the multimerization device during or after self-assembly.

13. The DNA sequence of any one of claims 1 to 12 encoding appropriately designed additional cysteines as part of the multimerization device or the linker(s) to allow for intermolecular covalent linkage to additional functional domains, which are separately produced in vivo or in vitro and coupled to the multimeric protein via oxidation of a free cystein to result in a covalent disulphide bridge.

14. The DNA sequence according to any one of claims 1 to 14 wherein said linker sequences b) and d) are the same.

15. The DNA sequence according to any one of claims 1 to 14 wherein said linker sequences b) and d) are different.

16. The DNA sequence according to any of claims 1-15 in which said linkers are derived from the human inter-domain linking sequence.

17. The DNA sequence according to claim 16 in which said human inter-domain linking sequence comprises at least a part of a human antibody hinge sequence.

18. The DNA sequence according to claim 17 in which said human inter-domain linking sequence comprises at least a part of a human IgG3 hinge sequence.

19. The DNA sequence according to any of claims 1-18 in which one or both of said functional domains either:
a) bind to a defined target substance, or
b) catalyze reaction of a defined substrate, or
c) inhibit the action of an enzyme, or
d) bind or block a receptor binding side, or
e) bind to a metal ion.

20. The DNA sequence according to claim 19 in which one or both of said functional domains comprise at least a part of a member of the immunoglobulin super-family.

21. The DNA sequence according to claim 20 in which one or both of said functional domains is a single chain Fv-fragment.

22. The DNA sequence according to claim 20 in which one or both of said members of the immunoglobulin super-family binds to a carbohydrate target substance.

23. The DNA sequence according to claim 22 in which one or both of said members of the immunoglobulin super-family binds to Lewis Y-related carbohydrate structures.

24. The DNA sequence according to any one of claims 1 to 23 wherein said multimerization device is derived from a randomized DNA library in a suitable host, from which DNA members are identified by their ability to encode multimerizing peptides.

25. A vector, preferably an expression vector, comprising a DNA sequence according to any of claims 1-24.

26. The vector according to claim 24 comprising a first DNA sequence encoding a signal sequence for bacterial secretion and a second DNA sequence according to any of claims 1-24.

27. A cellular host comprising at least one vector according to claim 25 or 26.

28. The cellular host according to claim 27 which is mammalian, preferably human, yeast, plant, insect, preferably Spodoptera frugiperda or bacterial, preferably an E. coli cell.

29. A protein, preferably a fusion protein, comprising:
a) a first functional domain,
b) a first linker sequence,

c) a preferably peptidic multimerization device of not more than 110 and preferably 30-80 amino acid residues in length, which is capable of self-assembly to a trimer or higher order oligomer, and which is of predominantly human composition,
d) a second linker sequence, and
e) a second functional domain.

30. A fusion protein encoded by the DNA sequence of any one of claims 1 to 24, the vector of any one of claims 25 to 26 or/and produced by the cellular host of claim 27 or 28.

31. A multimeric protein assembled from proteins and/or fusion proteins according to claim 29 or 30.

32. The multimeric protein of claim 31 which is a homomultimeric protein.

33. The multimeric protein of claim 31 which is a heteromultimeric protein.

34. A method of producing the fusion protein according to claim 29 or 30, said method comprising growing a cellular host according to claim 27 or 28 in an appropriate medium whereby said fusion protein is expressed, and isolating said fusion protein.

35. A method of producing a fusion protein according to claim 29 or 30, said method comprising growing a cellular host according to claim 27 or 28 in an appropriate medium whereby said protein is expressed, and secreted to the periplasm of said host, and

isolating said protein, preferably as a multimeric protein from said periplasm.

36. A method for producing a protein comprising
a) a first functional domain,
b) a first linker sequence,
c) a multimerization device of not more than 110 and preferably 30-80 amino acids in length, which is capable of self-assembly to a trimer or higher order oligomer, and which is of mammalian and predominantly human composition,
d) a second linker sequence, and
e) a second functional domain,
wherein at least one of a) to e) is
i) produced recombinantly, and, if all of a) to e) are produced recombinantly, at least two of the constituents are produced by different host cells; and/or
ii) produced synthetically; and/or
iii) produced semisynthetically; and/or
iv) produced by in-vitro translation; and
wherein at least two of a) to e) are combined by enzymatic and/or chemical coupling thereby giving rise to the complete protein and preferably to the complete multimeric protein.

37. A method of producing a multimeric protein, said method comprising associating at least three proteins according to claim 36, said method giving rise either to homomultimeric or heteromultimeric proteins.

38. A diagnostic composition comprising at least a fusion protein according to claim 29 or 30, and/or a

multimeric protein according to any one of claims 31 to 33.

39. A pharmaceutical composition comprising at least a fusion protein protein according to claim 29 or 30, and/or a multimeric protein according to any one of claims 31 to 33 optionally in combination with a pharmaceutically acceptable carrier.

40. A gene cassette comprising at least one DNA sequence according to any one of claims 1 to 24.

41. A kit comprising at least one gene cassette according to claim 40.

Fig.1A Modular cistron encoding a bifunctional, self-assembling fusion protein

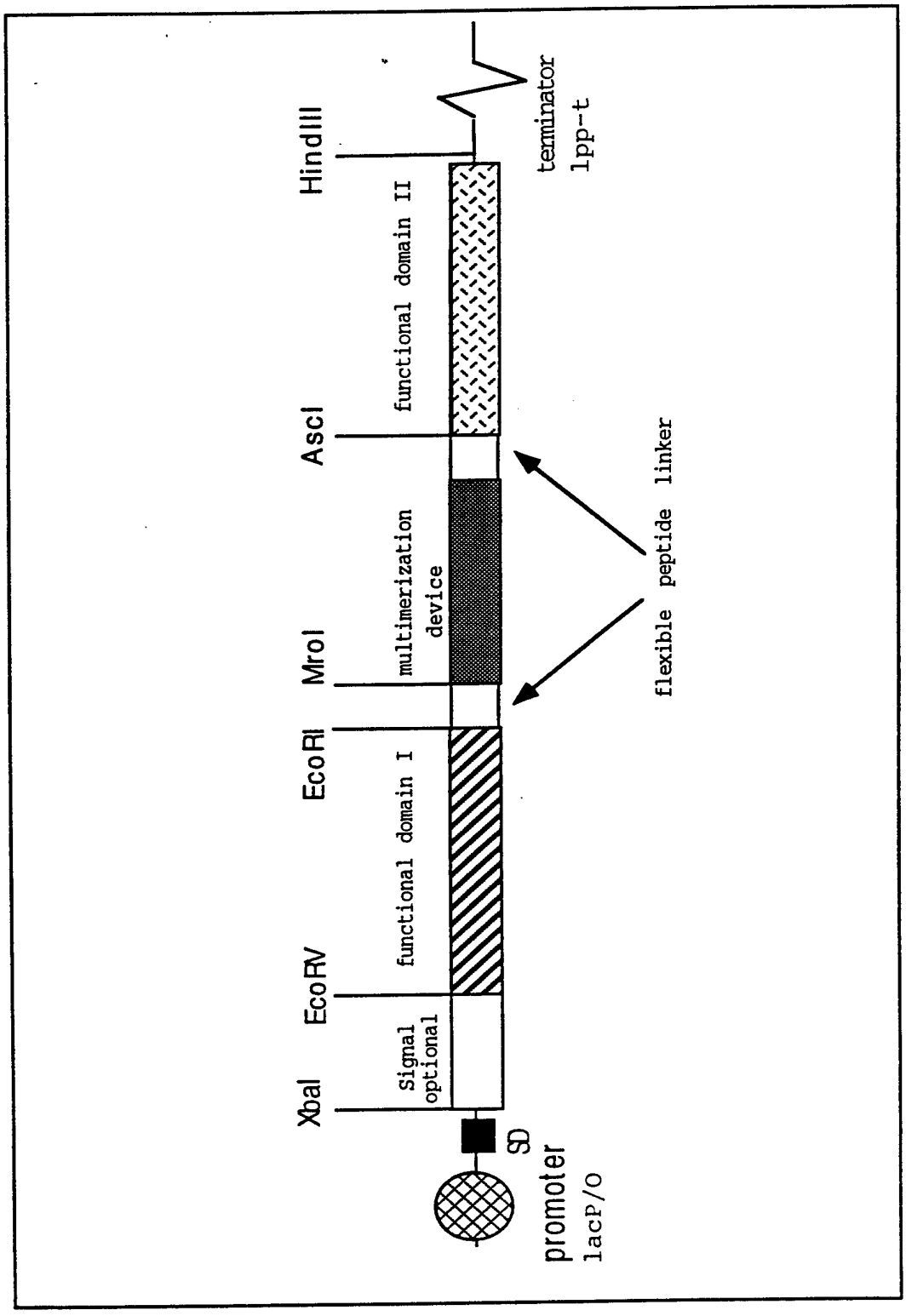


Fig. 1B. schematic illustration of the self-assembly process

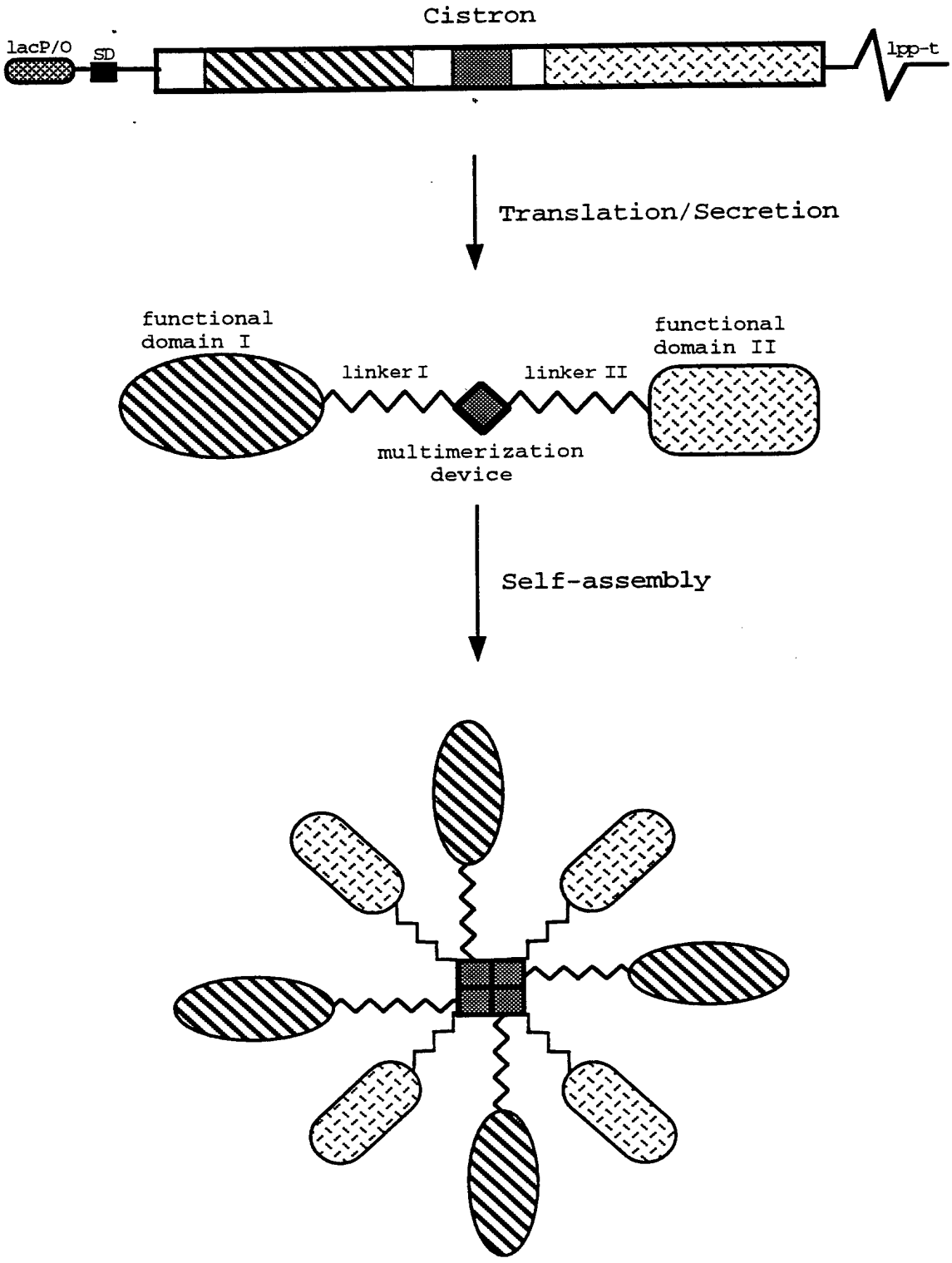


Fig.2: protein sequence of the tetramerization domain of human p53

```
         10          20          30          40
KPLDGEYFTL QIRGRERFEM FRELNEALEL KDAQAGKEPG G.
```

Fig.3 synthetic MroI-AscI gene cassette encoding for a multimerization device based on human p53 (note: the cleavage sequence of retriction sites in this and all following figures is simplified for illustration purposes, the synthetic inserts are sticky-ended with appropriate overhangs).

```
   MroI
    S   G   K   P   L   D   G   E   Y   F   T   L   Q   I   R   G   R   E
5' TCC GGA AAA CCA CTG GAT GGA GAA TAT TTC ACC CTT CAG ATC CGT GGG CGT GAG
           9           18          27          36          45          54
3' AGG CCT TTT GGT GAC CTA CCT CTT ATA AAG TGG GAA GTC TAG GCA CCC GCA CTC

    R   F   E   M   F   R   E   L   N   E   A   L   E   L   K   D   A   Q
    CGC TTC GAG ATG TTC CGA GAG CTG AAT GAG GCC TTG GAA CTC AAG GAT GCC CAG
             63          72          81          90          99         108
    GCG AAG CTC TAC AAG GCT CTC GAC TTA CTC CGG AAC CTT GAG TTC CTA CGG GTC

                                                    AscI
     A   G   K   E   P   G   G   S   G   G   A   P
    GCT GGG AAG GAG CCA GGG GGG AGC GGA GGC GCG CCG 3'
           117         126         135         144
    CGA CCC TTC CTC GGT CCC CCC TCG CCT CCG CGC GGC 5'
```

Fig.4: synthetic EcoRI-MroI gene cassette encoding a N-terminal flexible linker of human origin based on the upper hinge of human IgG3 (huIgG3) as a flexible, protease-stable peptide linker between a N-terminal functional domain and a peptidic multimerization device of this invention.

```
   EcoRI                                                MroI
   E   F   T   P   L   G   D   T   T   H   T   S   G
5' GAA TTC ACC CCG CTG GGT GAC ACC ACC CAC ACC TCC GGA  3'
           9          18          27          36
3' CTT AAG TGG GGC GAC CCA CTG TGG TGG GTG TGG AGG CCT  5'
```

Fig.5: AscI-HindIII gene cassette encoding a C-terminal metal binding domain

```
   AscI                                  Hind III
   A   P   H   H   H   H   H   H   *   *
5' GCG CCG CAC CAC CAC CAC CAC CAC TGA TAA GCT T 3'
           9          18          27
3' CGC GGC GTG GTG GTG GTG GTG CAC ACT ATT CGA A 5
```

Fig.6A. pACK phagemid vector for expression of the tetrameric and bifunctional fusion protein MSL5-p53-His in *E. coli*.

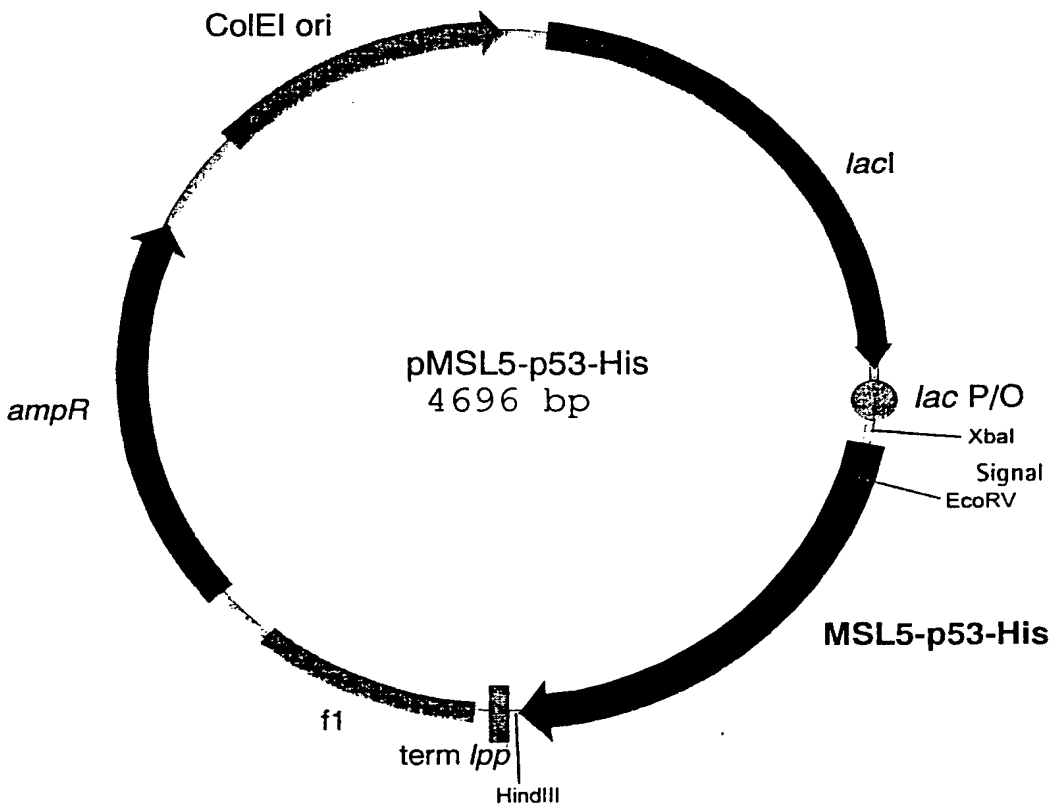


Fig. 6B.: Size exclusion chromatography of MSL-5 antibody constructs (•) (monomeric scFv, dimeric "miniantibody" and tetrameric/bifunctional MSL5-p53-His ("multibody")at 0.5 micromolar concentrations). The column (Superose 12, Pharmacia) was calibrated with standard proteins in HBS (o) (cytochrome c: 12,4 kD; carbonic anhydrase: 29 kD; albumin: 66 kD; alcohol dehydrogenase: 150 kD; ß-amylase: 200 kD).

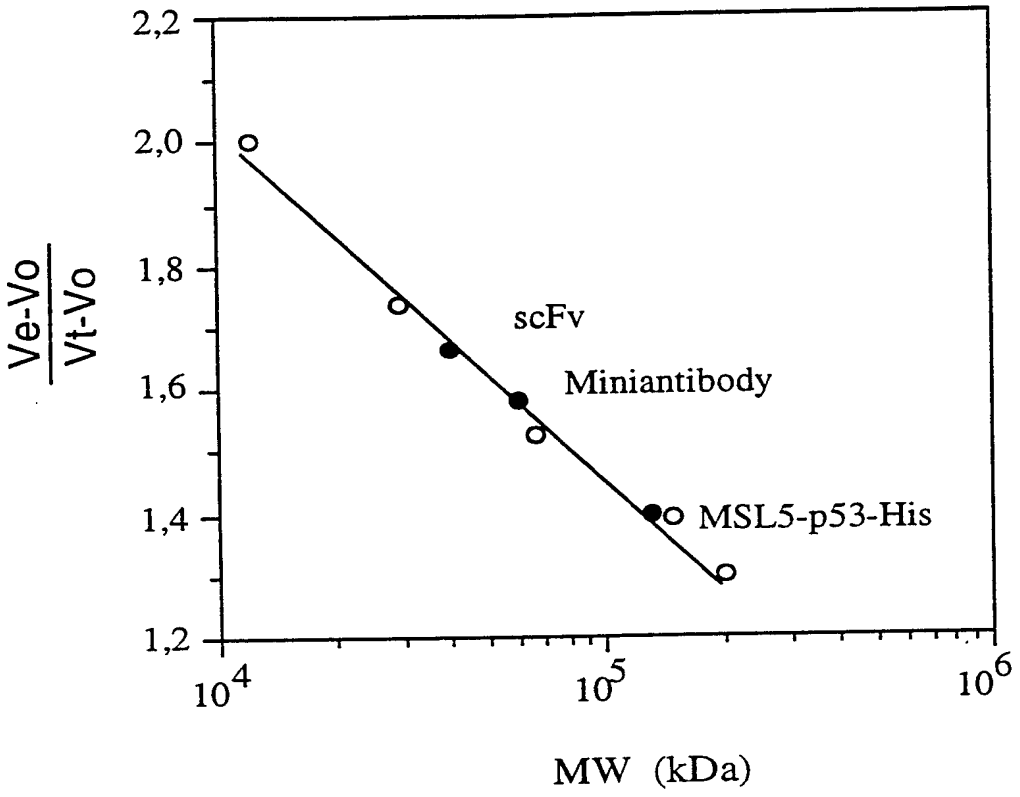


Fig 6C.: BIAcore overlay plot of individual dissociation curves of the MSL-5 IgM (14 nM) (A), the tetrameric and bifunctional protein MSL5-p53-His (1.07 µM) (B), the dimeric miniantibody (2.4 µM) (C) and monomeric scFv fragment (0.67 µM) (D).

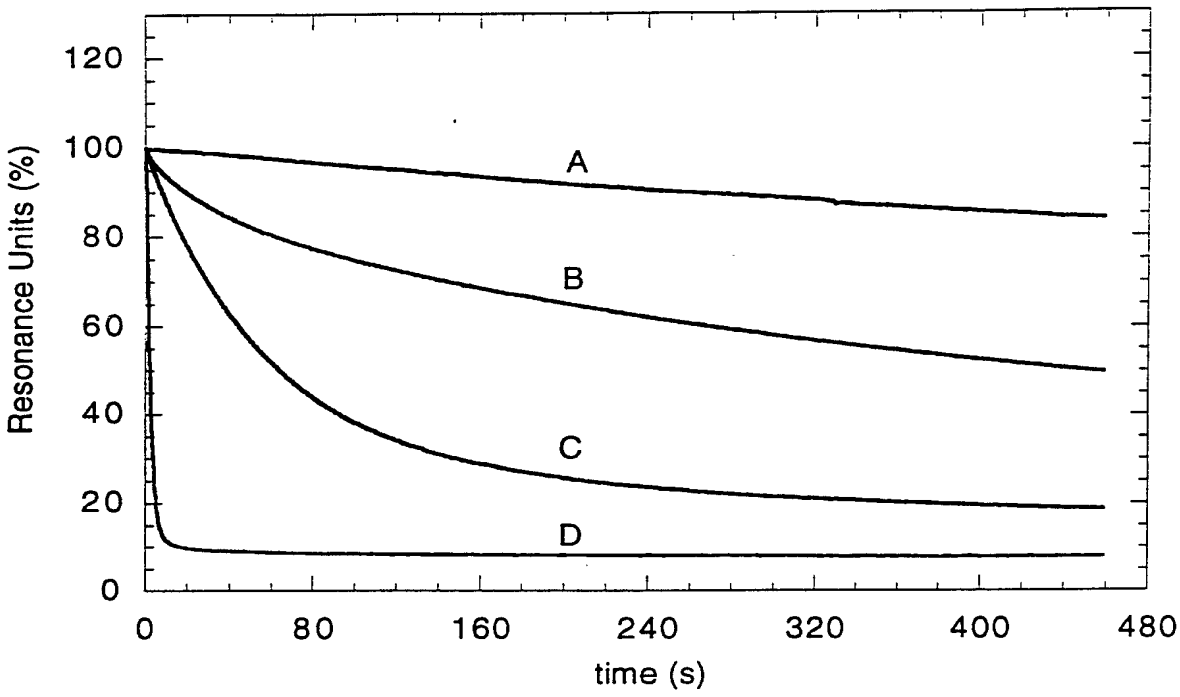


Fig.7 : Protein sequence of human Interleukin-2 (huIL2)

```
        10         20         30         40         50         60
APTSSSTKKT QLQLEHLLLD LQMILNGINN YKNPKLTRML TFKFYMPKKA TELKHLQCLE

        70         80         90        100        110        120
EELKPLEEVL NLAQSKNFHL RPRDLISNIN VIVLELKGSE TTFMCEYADE TATIVEFLNR

       130
WITFCQSIIS TLTDV*
```

Fig. 8.: Synthetic AscI-HindIII cassette encoding human IL2 as a C-terminal functional domain of this invention.

```
    AscI
      G   A   P   A   P   T   S   S   S   T   K   K   T   Q   L   Q   L   E
5'  GGC GCG CCG GCA CCT ACT TCA AGT TCT ACA AAG AAA ACA CAG CTA CAA CTG GAG
              9          18          27          36          45          54
3'  CCG CGC GGC CGT GGA TGA AGT TCA AGA TGT TTC TTT TGT GTC GAT GTT GAC CTC

      H   L   L   L   D   L   Q   M   I   L   N   G   I   N   N   Y   K   N
      CAT TTA CTG CTG GAT TTA CAG ATG ATT TTG AAT GGA ATT AAT AAT TAC AAG AAT
               63          72          81          90          99         108
      GTA AAT GAC GAC CTA AAT GTC TAC TAA AAC TTA CCT TAA TTA TTA ATG TTC TTA

      P   K   L   T   R   M   L   T   F   K   F   Y   M   P   K   K   A   T
      CCC AAA CTC ACC AGG ATG CTC ACA TTT AAG TTT TAC ATG CCC AAG AAG GCC ACA
             117         126         135         144         153         162
      GGG TTT GAG TGG TCC TAC GAG TGT AAA TTC AAA ATG TAC GGG TTC TTC CGG TGT

      E   L   K   H   L   Q   C   L   E   E   E   L   K   P   L   E   E   V
      GAA CTG AAA CAT CTT CAG TGT CTA GAA GAA GAA CTC AAA CCT CTG GAG GAA GTG
             171         180         189         198         207         216
      CTT GAC TTT GTA GAA GTC ACA GAT CTT CTT CTT GAG TTT GGA GAC CTC CTT CAC

      L   N   L   A   Q   S   K   N   F   H   L   R   P   R   D   L   I   S
      CTA AAT TTA GCT CAA AGC AAA AAC TTT CAC TTA AGA CCC AGG GAC TTA ATC AGC
             225         234         243         252         261         270
      GAT TTA AAT CGA GTT TCG TTT TTG AAA GTG AAT TCT GGG TCC CTG AAT TAG TCG


      N   I   N   V   I   V   L   E   L   K   G   S   E   T   T   F   M   C
      AAT ATC AAC GTA ATA GTT CTG GAA CTA AAG GGA TCT GAA ACA ACA TTC ATG TGT
             279         288         297         306         315         324
      TTA TAG TTG CAT TAT CAA GAC CTT GAT TTC CCT AGA CTT TGT TGT AAG TAC ACA

      E   Y   A   D   E   T   A   T   I   V   E   F   L   N   R   W   I   T
      GAA TAT GCT GAT GAG ACA GCA ACC ATT GTA GAA TTT CTG AAC AGA TGG ATT ACC
             333         342         351         360         369         378
      CTT ATA CGA CTA CTC TGT CGT TGG TAA CAT CTT AAA GAC TTG TCT ACC TAA TGG

                                                              HindIII
      F   C   Q   S   I   I   S   T   L   T   D   V   *   *
      TTT TGT CAA AGC ATC ATC TCA ACA CTG ACT GAC GTC TGA TAA GCT T-3'
             387         396         405         414
      AAA ACA GTT TCG TAG TAG AGT TGT GAC TGA CTG CAG ACT ATT CGA A-5'
```

Fig. 9. CTLL-A proliferation assay for the determation of IL2-activity of the bifunctional and multimeric protein H11-p53-huIL2.

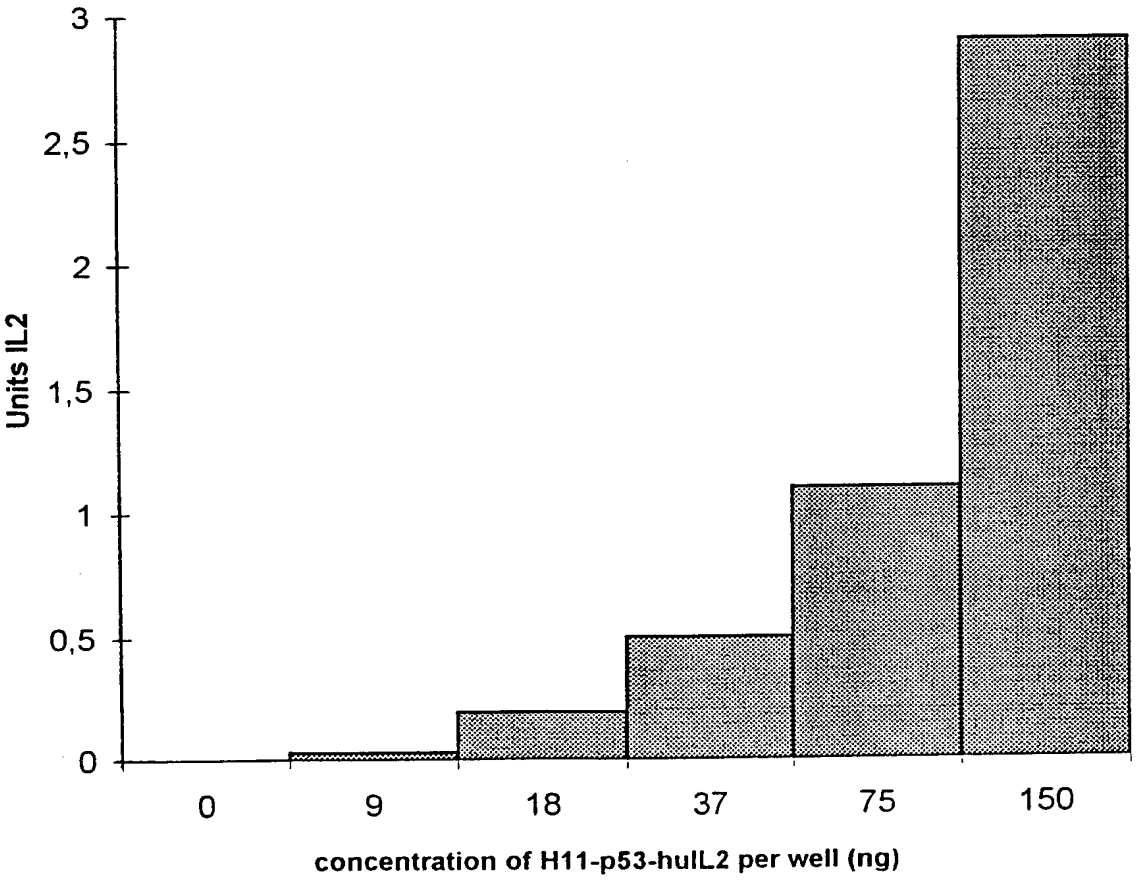


Fig. 10. Anti-ESL-1 ELISA: measurable affinity (avidity) of 12 library-derived tetrameric and bifunctional scFv-p53-His proteins to ESL-1 in comparison to the monomeric scFv-His formats. Binding proteins are detected with antibodies against the C-terminal functional domain, the polyhistidine metal binding site.

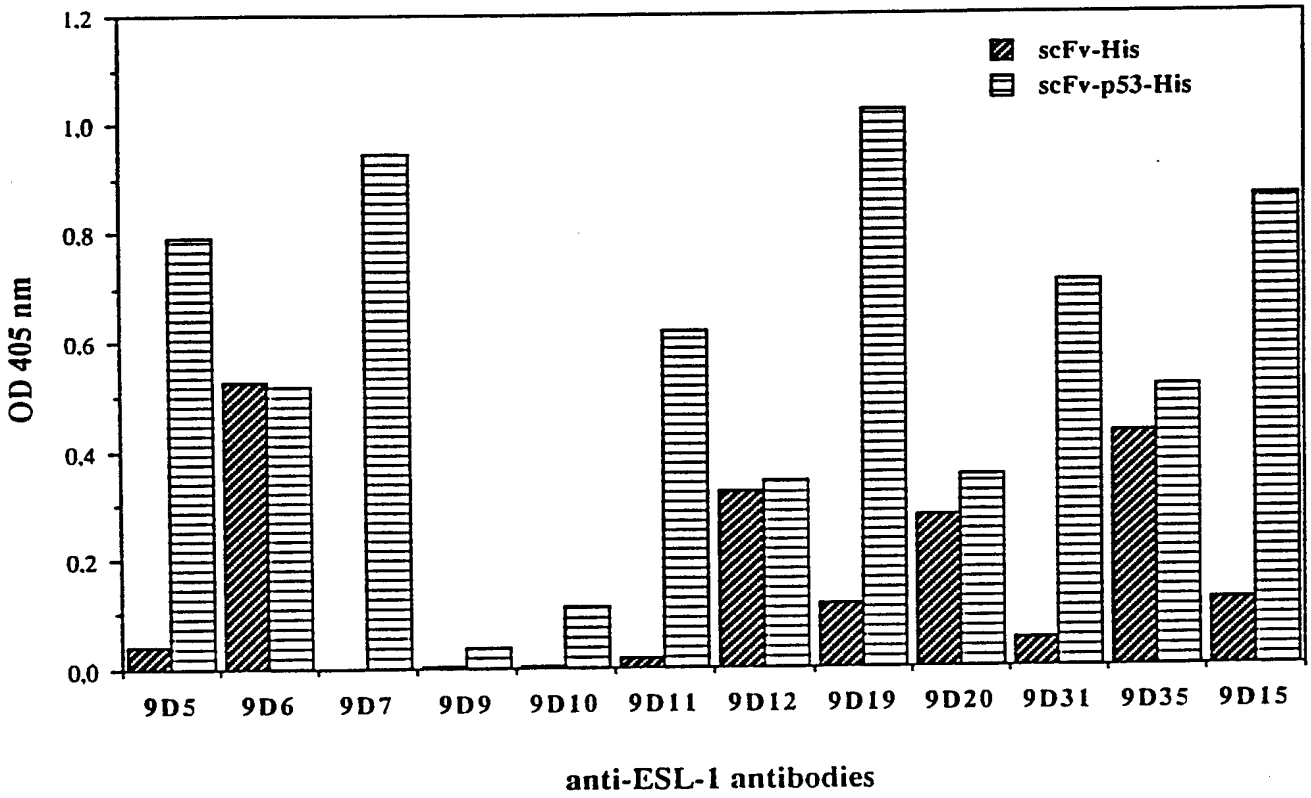


Fig. 11a.: Residues 13 to 87 of the human $hTAF_{II}31$ as a hetero-tetramerization device of this invention

P K D A Q M M A Q I L K D M G I T E Y E P R V I N Q M L E F
A F R Y V T T I L D D A K I Y S S H A K K A T V D A D D V R
L A I Q C R A D Q S F T S P P.

Fig. 11b.: Residues 10 to 82 of the human $hTAF_{II}80$ as a hetero-tetramerization device of this invention

S N T V L P S E S M K V V A E S M G I A Q I Q E E T C Q L L
T D E V S Y R I K E I A Q D A L K F M H M G K R Q K L T T S
D I D Y A L K L K N V E P

Fig. 12a.: residues 67 to 134 of the human histone 3 (H3) as a hetero-tetramerization device of this invention

P F Q R L M R E I A Q D F K T D L R F Q S S A V A L Q E A C
E S Y L V G L F E D T N L C V I H A K R V T I M P K D I Q L
A R R I R G

Fig 12b.: Residues 29 to 95 of the human histone 4 (H4) as a hetero-tetramerization device of this invention

G I T K P A I R R L A R R G G V K R I S G L I Y E E T R G V
L K V F L E N V I R D A V T Y T E H A K R K T V T A M D V V
Y A L K R Q G

Fig.13: Protein sequence of the platelet aggregation inhibitor decorsin as a functional domain of this invention.

```
1          11         21         31
APRLPQCQGD DQEKCLCNKD ECPPGQCRFP RGDADPYCE
```

Fig. 14.: Synthetic EcoRV-EcoRI cassette encoding the platelet aggregation inhbitor decorsin as a N-terminal functional domain of this invention.

```
EcoRV
D   I   A   P   R   L   P   Q   C   Q   G   D   D   Q   E   K   C   L
GAT ATC GCT CCG CGT CTG CCG CAG TGC CAG GGT GAC GAC CAG GAA AAA TGC CTG
          9          18          27          36          45          54
CTA TAG CGA GGC GCA GAC GGC GTC ACG GTC CCA CTG CTG GTC CTT TTT ACG GAC

C   N   K   D   E   C   P   P   G   Q   C   R   F   P   R   G   D   A
TGC AAC AAA GAC GAA TGC CCG CCG GGT CAG TGC CGT TTC CCG CGT GGT GAC GCT
         63          72          81          90          99         108
ACG TTG TTT CTG CTT ACG GGC GGC CCA GTC ACG GCA AAG GGC GCA CCA CTG CGA

                EcoRI
D   P   Y   C   E   F
GAC CCG TAC TGC GAA TTC  3'
        117         126
CTG GGC ATG ACG CTT AAG  5'
```

Fig.15: Synthetic AscI-HindIII cassette encoding the platelet aggregation inhibitor decorsin as a C-terminal functional domain of this invention.

```
AscI
A   P   A   P   R   L   P   Q   C   Q   G   D   D   Q   E   K   C   L
GCG CCG GCT CCG CGT CTG CCG CAG TGC CAG GGT GAC GAC CAG GAA AAA TGC CTG
         12          21          30          39          48          57
CGC GGC CGA GGC GCA GAC GGC GTC ACG GTC CCA CTG CTG GTC CTT TTT ACG GAC

C   N   K   D   E   C   P   P   G   Q   C   R   F   P   R   G   D   A
TGC AAC AAA GAC GAA TGC CCG CCG GGT CAG TGC CGT TTC CCG CGT GGT GAC GCT
         66          75          84          93         102         111
ACG TTG TTT CTG CTT ACG GGC GGC CCA GTC ACG GCA AAG GGC GCA CCA CTG CGA

                        HindIII
D   P   Y   C   E   *   *
GAC CCG TAC TGC GAA TGA TAA GCT T 3'
        120         129
CTG GGC ATG ACG CTT ACT ATT CGA A 5'
```

Fig.16: Peptide sequence of an endotoxin-neutralizing peptide as a functional domain of this invention

```
          1          11
          RWKVRKSFFKL Q
```

Fig.17: Synthetic EcoRV-EcoRI cassette encoding an endotoxin-neutralizing peptide as a N-terminal functional domain of this invention.

```
EcoRV                                                    EcoRI
   I   M   R   W   K   V   R   K   S   F   F   K   L   Q   E   F
5' ATC ATG CGT TGG AAA GTT CGT AAA TCT TTC TTC AAA CTG CAG GAA TTC 3'
             9          18          27          36          45
3' TAG TAC GCA ACC TTT CAA GCA TTT AGA AAG AAG TTT GAC GTC CTT AAG 5
```

Fig.18: Synthetic AscI-HindIII cassette encoding an endotoxin-neutralizing peptide as a C-terminal functional domain of this invention.

```
AscI                                                        HindIII
   A   P   R   W   K   V   R   K   S   F   F   K   L   Q   *   *
5' GCG CCG CGT TGG AAA GTT CGT AAA TCT TTC TTC AAA CTG CAG TGA TAAGCTT  3'
             9          18          27          36          45
3' CGC GGC GCA ACC TTT CAA GCA TTT AGA AAG AAG TTT GAC GTC ACT ATTCGAA  5'
```

Fig.19: protein sequence of human plasma protein PF4

```
 1 mssaagfcas rpgllflgll llplvvafas aeaeedgdlq
41 clcvkttsqv rprhitslev ikagphcpta qliatlkngr
81 kicldlqapl ykkiikklle s
```

Fig. 20: Synthetic MroI-AscI-HindIII gene cassette encoding a a tetramerization device of this invention based on the human plasma protein PF4.

```
    MroI
     S   G   M   S   S   A   A   G   F   C   A   S   R   P   G   L   L   F
5'  TCC GGA ATG TCT TCT GCT GCT GGT TTC TGC GCT TCT CGT CCG GGT CTG CTG TTC
              9          18          27          36          45          54
3'  AGG CCT TAC AGA AGA CGA CGA CCA AAG ACG CGA AGA GCA GGC CCA GAC GAC AAG

    L   G   L   L   L   L   P   L   V   V   A   F   A   S   A   E   A   E
    CTG GGT CTG CTG CTG CTG CCG CTG GTT GTT GCT TTC GCT TCT GCT GAA GCT GAA
             63          72          81          90          99         108
    GAC CCA GAC GAC GAC GAC GGC GAC CAA CAA CGA AAG CGA AGA CGA CTT CGA CTT

    E   D   G   D   L   Q   C   L   C   V   K   T   T   S   Q   V   R   P
    GAA GAC GGT GAC CTG CAG TGC CTG TGC GTT AAA ACC ACC TCT CAG GTT CGT CCG
            117         126         135         144         153         162
    CTT CTG CCA CTG GAC GTC ACG GAC ACG CAA TTT TGG TGG AGA GTC CAA GCA GGC

    R   H   I   T   S   L   E   V   I   K   A   G   P   H   C   P   T   A
    CGT CAC ATC ACC TCT CTG GAA GTT ATC AAA GCT GGT CCG CAC TGC CCG ACC GCT
            171         180         189         198         207         216
    GCA GTG TAG TGG AGA GAC CTT CAA TAG TTT CGA CCA GGC GTG ACG GGC TGG CGA

    Q   L   I   A   T   L   K   N   G   R   K   I   C   L   D   L   Q   A
    CAG CTG ATC GCT ACC CTG AAA AAC GGT CGT AAA ATC TGC CTG GAC CTG CAG GCT
            225         234         243         252         261         270
    GTC GAC TAG CGA TGG GAC TTT TTG CCA GCA TTT TAG ACG GAC CTG GAC GTC CGA

    P   L   Y   K   K   I   I   K   K   L   L   E   S   G   G   S   G   G
    CCG CTG TAC AAA AAA ATC ATC AAA AAA CTG CTG GAA TCT GGG GGG AGC GGA GGC
            279         288         297         306         315         324
    GGC GAC ATG TTT TTT TAG TAG TTT TTT GAC GAC CTT AGA CCC CCC TCG CCT CCG


      AscI
    A   P
    GCG CCG 3'
            333
    CGC GGC 5'
```

Fig. 21: Protein sequence of residues 209 to 273 of the human TSP4

209 EP LAATGTGDFN RQFLGQMTQL NQLLGEVKDL LRQQVKETSF

251 LRNTIAECQA CGPLKFQSPT PST

Fig. 22: Synthetic MroI-AscI gene cassette encoding a pentamerization device of this invention based on residues 209 to 273 of the human TSP4.

```
    MroI
    S   G   E   P   L   A   A   T   G   T   G   D   F   N   R   Q   F   L
5'  TCC GGA GAA CCG CTG GCG GCG ACC GGC ACC GGC GAT TTT AAC CGT CAG TTT CTG
              9          18          27          36          45          54
3'  AGG CCT CTT GGC GAC CGC CGC TGG CCG TGG CCG CTA AAA TTG GCA GTC AAA GAC

    G   Q   M   T   Q   L   N   Q   L   L   G   E   V   K   D   L   L   R
    GGC CAG ATG ACC CAG CTG AAC CAG CTG CTG GGC GAA GTG AAA GAT CTG CTG CGT
             63          72          81          90          99         108
    CCG GTC TAC TGG GTC GAC TTG GTC GAC GAC CCG CTT CAC TTT CTA GAC GAC GCA

    Q   Q   V   K   E   T   S   F   L   R   N   T   I   A   E   C   Q   A
    CAG CAG GTG AAA GAA ACC AGC TTT CTG CGT AAC ACC ATT GCG GAA TGC CAG GCG
            117         126         135         144         153         162
    GTC GTC CAC TTT CTT TGG TCG AAA GAC GCA TTG TGG TAA CGC CTT ACG GTC CGC

     C   G   P   L   K   F   Q   S   P   T   P   S   T   G   G   S   G   G
    TGC GGC CCG CTG AAA TTT CAG AGC CCG ACC CCG AGC ACC GGG GGG AGC GGA GGC
            171         180         189         198         207         216
    ACG CCG GGC GAC TTT AAA GTC TCG GGC TGG GGC TCG TGG CCC CCC TCG CCT CCG

    AscI
    A   P
    GCG CCG  3'
        222
    CGC GGC  5'
```

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/10, 15/12, 15/24, 15/62, 15/70

A1

(11) International Publication Number: **WO 97/32017**

(43) International Publication Date: 4 September 1997 (04.09.97)

(21) International Application Number: PCT/EP97/00931

(22) International Filing Date: 26 February 1997 (26.02.97)

(30) Priority Data:
96102852.9 26 February 1996 (26.02.96) EP
(34) Countries for which the regional or international application was filed: DE et al.

(71) Applicant *(for all designated States except US)*: MORPHOSYS GESELLSCHAFT FÜR PROTEINOPTIMIERUNG MBH [DE/DE]; Frankfurter Ring 193a, D-80807 München (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: ILAG, Vic [PH/DE]; Knorrstrasse 85, D-80807 München (DE). GE, Liming [CN/DE]; Portiastrasse 12, D-81545 München (DE).

(74) Agent: VOSSIUS & PARTNER; P.O. Box 86 07 67, D-81634 München (DE).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

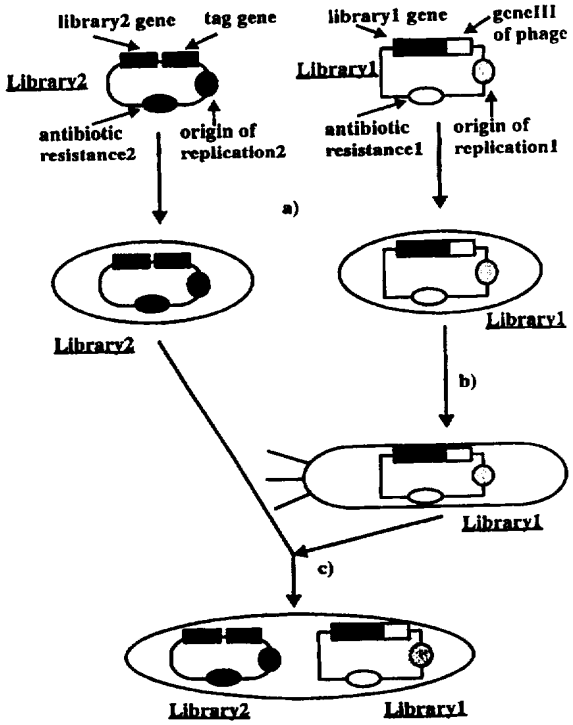
Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NOVEL METHOD FOR THE IDENTIFICATION OF NUCLEIC ACID SEQUENCES ENCODING TWO OR MORE INTERACTING (POLY)PEPTIDES

(57) Abstract

The present invention relates to methods for identifying nucleic acid sequences which encode two or more specific interacting peptides or proteins. Furthermore, the present invention relates to kits which may be used for the identification of nucleic acid sequences in accordance with the method of the present invention.

General description of the polyphage principle



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia
AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

NOVEL METHOD FOR THE IDENTIFICATION OF NUCLEIC ACID SEQUENCES ENCODING TWO OR MORE INTERACTING (POLY)PEPTIDES

The present invention relates to methods for identifying nucleic acid sequences which encode two or more specific interacting peptides or proteins. Furthermore, the present invention relates to kits which may be used for the identification of nucleic acid sequences in accordance with the method of the present invention.

Protein-protein interactions play an important role in all biological processes, from the replication and expression of genes to the morphogenesis of organisms (Lewin, B. 1994, Genes V. Oxford University Press). Methods for detecting protein-protein interactions have proved useful in understanding the basic mechanisms of different biological processes and the development of therapeutics. Detection of protein-protein interactions can be divided into two main categories: (i) physico-chemical based and (ii) genetic approaches (Phizicky, E.,M. & Fields, S. Microbiological Reviews 59 (1995) 94-123). Detection of protein-protein interactions by physico-chemical methods usually requires significant amounts of material, and more importantly, the identity of the proteins to be studied must be known. Recent developments in methods of mass spectrometry circumvent this problem but such suffer the disadvantage of requiring sophisticated equipment and expertise (Wang, R. & Chait, B.T., Current Opinion in Biotech. 5 (1994) 77-84). In contrast, genetic approaches provide an easy and powerful method of identifying protein-protein interactions without the need for pure material and specialized equipment, with the added advantage of higher throughput.

Different genetic approaches have been used to identify protein-protein interactions. The current method of choice is the yeast 2-hybrid system (Fields, S. & Song, O.K.,

Nature (London) 340, (1989) 245-246) which allows the identification of novel proteins that interact with a known protein.

Another popular genetic approach is the phage display system (Patent Application WO90/02809) whereby proteins are fused to a component of a surface protein of filamentous phage to allow selection for binding to a ligand of interest. The gene encoding the protein displayed on the surface of the phage is packaged inside the phage allowing the coupling of genetic information with the gene product. This allows the screening of "libraries" of proteins whereby the identity of the screened protein is deduced from the nucleic acid sequence of the phage. This technique has been extended by Winter et al. (Patent Application WO 92/20791) to produce libraries of multimeric members of a specific binding pair (e.g. combinations of VH and VL chains of an antibody) and select for functional specific binding pair members that can bind to the complementary specific binding pair member (e.g. antigen). Said libraries are constructed by combining two sub-libraries each encoding a collection of corresponding sub-units of said multimeric members (e.g. a library of VH chains is combined with a library of VL chains) wherein in principle each sub-unit out of the first sub-library is able to bind to each sub-unit out of the second sub-library non-specifically. Although this method has led to the identification of unique antibodies against particular antigens, it fails to provide a method for identifying two partners of a specific binding pair when both are unknown.

A unique version of phage display which relies on non-infective phage has recently been proposed (Duenas, M. & Borrebaeck, C. A. K., Bio/Technology 12 (1994) 999-1002; EP 0 614 989). A version of this system that led to the identification of proteins from a cDNA library that interacts with the jun protein has been described (Gramatikoff et al., Nucleic. Acids Res. 22 (1994) 5761-5762). The same principle has been also shown to work with an antibody-antigen system (Krebber et al., FEBS Letters 377 (1995) 227-231).

In spite of the power of all the aforementioned genetic selection approaches, they are limited to the selection of interacting binding entities from only a single genetically-diverse population (library vs. individual).

It would, however, be highly desirable to simultaneously identify binding entities and their specific binding partners in a library vs. library setting, wherein preferably at least two genetically diverse populations are involved. A solution to this technical problem, i.e. the identification of interacting entities and the respective nucleic acid sequences from more than one genetically diverse population (library vs. library) is neither provided nor suggested by the prior art. The present invention solves the above technical problem by providing the embodiments characterized in the claims. By using these embodiments, it has become possible to increase exponentially the rate at which (poly)peptide-(poly)peptide interactions are detected. The present invention may find applications in the field of functional genomics, whereby different proteins of unknown functions can be related with other proteins.

Accordingly, the present invention relates to a method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:

(a) providing a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;

(b) providing a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as mentioned in step (a), wherein the vector molecules

employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (a) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (a) and (b), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;

(c) optionally, providing additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (a) and/or step (b), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (a) and (b) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (c) together with at least one of said properties displayed by either said vector molecule and/or said recombinant insert used in steps (a) and/or (b), upon the interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;

(d) expressing members of said libraries of recombinant vectors or nucleic acid sequences mentioned in steps (a), (b) and optionally (c), in appropriate host cells so that at least one interaction is established;

(e) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;

(f) optionally, carrying out further selection, screening and/or purification steps; and

(g) identifying said nucleic acid sequences encoding said (poly)peptides.

Thus, in the context of the present invention, the term "properties that are phenotypically distinguishable" relates alternatively to properties that are encoded by the vector molecule or to properties that are encoded by the recombinant insert or to both types of properties. As regards the vector-encoded properties, these may e.g. be resistance markers or requirements for special nutrients. It should be noted that the recombinant insert may comprise a nucleic acid portion encoding said property in addition to the nucleic acid portion responsible for the interaction.

In the context of the present invention, the term "different member " denotes a different entity which may be, but is not necessarily, structurally different.

Further, in the context of the present invention, the term "plurality" bears the meaning of "at least two".

The novel properties generated by the at least two recombinant inserts reflect the inventive principle of the present invention. That is, only if two (or more) (poly)peptides interact, for example, in a homo-dimeric or hetero-dimeric fashion, a screenable or selectable property is generated. The interaction between the two or more molecules may be a direct one or may be mediated indirectly. Examples for a direct interaction are the binding of an antibody encoded by a nucleic acid sequence from library 1 to a cDNA protein from library 2, the binding of a protein encoded by a nucleic acid sequence from cDNA library 1 to a protein from a cDNA library 2, as well as of an anti-idiotypic antibody encoded by a nucleic acid sequence from one of the libraries to a corresponding antibody encoded by a nucleic acid sequence from the other library. The nucleic acid sequences are preferably DNA and most preferably genes or parts thereof.

An example of an indirect interaction is the bridging of two (poly)peptides encoded by the two libraries which is mediated by a phosphorylating enzyme. Once the phosphorylation of one (poly)peptide encoded e.g. by library 1 is effected by the respective kinase, then this protein is capable of interacting with the second (poly)peptide encoded by library 2. The phosphorylating enzyme exemplifying this type of interaction may be encoded by a nucleic acid from (one of) the additional libraries and/or may be encoded by the genome of the host cell. Typically, the interaction of the two (poly)peptides forms a "bridge" of molecules, said "bridge" being detectable using an appropriate detection process. Conveniently, said bridge is detectable by a tag molecule that is associated with, encoded by or attached to one of the (poly)peptides encoded by library 1 or preferably 2.

Furthermore, the present invention relates to a method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:

(a) expressing in appropriate host cells

(aa) nucleic acid sequences contained in a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;

(ab) nucleic acid sequences contained in a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as

mentioned in step (aa), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (aa) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (aa) and (ab), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;

(ac) optionally, nucleic acid sequences contained in additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (aa) and/or step (ab), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (aa) and (ab) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (ac) together with at least one of said properties displayed by either said vector molecule and/or said recombinant inserts used in steps (aa) and/or (ab), upon the interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;

so that at least one interaction is established;

(b) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;

(c) optionally, carrying out further screening, selection and/or purification steps; and

(d) identifying said nucleic acid sequences encoding said (poly)peptides.

In a preferred embodiment of the method of the present invention, said screenable or selectable property is expressed extracellularly.

This embodiment is conveniently employed in a number of laboratories which would make use of rather conventional methodology of the extracellular detection of such properties, e.g. by column chromatography wherein the e.g. screenable tag is retained, in combination with e.g. plaque purification techniques, which allow the further purification of the cells that were originally enriched by e.g. the column chromatography step.

In a further preferred embodiment of the method of the present invention, said recombinant vector molecule in step (a)/(aa) (the step identified after the slash refers to the corresponding step of the second embodiment of the method of the invention identified hereinabove) gives rise to a replicable genetic package (RGP) displaying said (poly)peptides at its surface. In this context, the term replicable genetic package (RGP) refers to an entity, such as a virus or bacteriophage, which can be replicated following infection of a suitable host cell. In the case of bacteriophage, for example, the collection of nucleic acid sequences can be inserted into either a phage or phagemid vector in frame with a component of the phage coat, such as gene III, resulting in display of the encoded binding entities on the surface of the phage. Particularly preferred as a

recombinant vector molecule is a recombinant phage, phagemid or virus, wherein said phage is most preferably

(a) one of the class I phage fd, M13, If, Ike, ZJ/2, Ff;

(b) one of the class II phage Xf, Pf1, and Pf3;

(c) one of the lambdoid phages, lamda, 434, P1;

(d) one of the class of enveloped phages, PRD1; or

(e) one of the class paramyxo-viruses, orthomyxo-viruses, baculo-viruses, retro-viruses, reo-viruses and alpha-viruses.

In a further preferred embodiment of the method according to the invention, said selection step (e)/(b) is carried out by selecting polyphage comprising the interacting (poly)peptides. Polyphage contain more than one copy of phage genomic DNA. They occur naturally at a low to moderate frequency when a newly forming phage coat encapsulates two or more single-stranded DNA molecules. In the case of the present invention, the polyphage which are formed will contain at least two phage genomes, which may either (i) both be representatives of library 1, or (ii) both be representatives of library 2, or (iii) be representatives of each of library 1 and library 2, or (iv) be a combination of (i) to (iii) with at least one member of one of the additional libraries. The efficiency of polyphage production can be increased by the introduction of appropriate mutations into the phage genome, as is well known to those skilled in the art (see, for example, Lopez, J. and Webster, R.E., Virology 127 (1983), 177-193, Bauer, M. and Smith, G.P., Virology 167 (1988) 166-175, or Gailus, V. et al., Res. Microbiol. 145 (1994) 699-709).

In a further preferred embodiment of the method of the invention, said screenable or selectable property is connected to the infectivity of said RGP.

In this embodiment, use is made of the possibility that the infectivity of e.g. a bacteriophage can be manipulated, said infectivity being directly correlated with the interaction of said (poly)peptides.

In a most preferred embodiment of the method of the present invention, said RGP is encoded by said recombinant vector used in step (a)/(aa) and rendered non-infective and infectivity of said RGP is restored by interaction of said (poly)peptide of step (a)/(aa) with the (poly)peptide of step (b)/(ab) and/or (c)/(ac), said (poly)peptide of step (b)/(ab) and/or (c)/(ac) being fused to a domain that confers infectivity to said RGP.

In a further most preferred embodiment of the method of the invention, said RGP is rendered non-infective by modification of a genetic sequence which encodes a surface protein necessary for the RGP's binding to and infection of a host cell.

These preferred and most preferred embodiments of the method of the present invention relating to the infectivity of the RGP serve as an alternative to the use of the screenable tag. In these embodiments, advantage can be taken of the phenomenon of selective infection (Krebber et al., FEBS Letters 377 (1995) 227-239). While the screenable tag enables physical separation of molecules from others in the population, the use of selective infection enables positive selection for the interacting pair. This phenomenon relies on the use of a construct which can selectively restore infectivity to phage which have been rendered non-infective by, for example, deletion of all but the C-terminus of the gene III protein. Use of such phage for displaying library 1 gives non-infectious phage carrying the binding entity. Co-expression with library 2 allows interactions between binding entities and binding partners to be established, as described above. Although the phage which carry the binding entity-binding partner pair are non-infective, infectivity can be restored if, in place of the screenable tag referred to above, an infectivity protein is used. In this context, the term infectivity protein refers to a substance which, when associated with the phage, can enable it to penetrate a bacterial host, where it is subsequently replicated. An example of an infectivity protein is the N-terminus (at least the first 220 amino acids) of gene III protein of the filamentous bacteriophage.

The infectivity protein confers on those phage which carry it, the ability to be replicated. Thus, only those phage which carry the binding entity/partner pair are replicated. Purification of hybrid phage containing genes from both libraries 1 and 2 then relies e.g. on the use of two selectable markers as indicated above. The genes in the phage can then be identified using methodology well known to those skilled in the art.

An additional preferred embodiment of the present invention relates to a method, wherein said recombinant vector molecules in step (a)/(aa) give rise to a fusion protein which is expressed on the surface of a cell, preferably a bacterium.

These fusion proteins, upon interaction with a suitable binding partner from library 2 connected e.g. with a screenable tag can be detected on the surface of host cells which may be, for example, bacteria, yeast, insect cells or mammalian cells. The display of fusion proteins on bacterial surfaces per se is well known in the art. Thus, lipoproteins (Lpp), outer membrane proteins A (OmpA), and flagella have been used to target antibodies and peptides to the cell surface of E.coli. Fuchs et al., Bio/Technology 9 (1991) 1369-1372, WO93/01287, presented a single chain antibody on the surface of E.coli as a fusion protein with the N-terminus of the peptidoglycan-associated lipoprotein. The antibody was visualized by the binding of fluorescently labeled antigen and fluorescently labeled antibodies directed to the linker peptide of the displayed single chain antibody. Francisco et al., Proc. Natl. Acad. Sci. USA 90 (1993) 10444-10448, and Georgiu, G. et al., WO93/10214, displayed antibodies on the E.coli surface by fusing the N-terminus of a single chain antibody to the C-terminus of OmpA while the N-terminus of OmpA was fused to the signal sequence and the first nine amino acids of Lpp. Binding of a fluorescently labeled antigen to the OmpA-antibody fusion protein was detected by FACS. Klauser (WO 95/17509) transferred the IgA protease system from Neisseria to E.coli to facilitate display of antibodies. Integration of the beta-domain of the IgA protease precursor into the outer membrane lead to the transport of the

protease domain across the membrane followed by autoproteolytic release into the medium. Antibodies linked to the beta-domain of IgA protease are therefore presented on the surface of bacteria. Further, Lu, Z. et al., Bio/Technology 13 (1994) 366-371, described a system for displaying peptides on the surface of the bacterium by fusing it to thioredoxin and the bacterial flagella, to screen for peptide mimics of the epitope for an anti-IL-8 antibody.

The further identification of the desired nucleic acid molecule encoding the interacting (poly)peptides may then be effected by methods known in the art, e.g. by purifying host cells displaying a tag on their surface and further by antibioticum-based selection techniques, DNA purification and sequencing.

In a particularly preferred embodiment of the method of the present invention, said bacterium is Neisseria gonorrhoe or E.coli and said fusion protein consists of at least a part of a flagellum, lam B, peptidoglycan-associated lipoprotein or the Omp A protein and said (poly)peptide.

As has been repeatedly pointed out hereinabove, a tag connected to the (poly)peptide encoded by library 2 can conveniently be used in the identification strategy of the desired nucleic acid sequences. Accordingly, in a further preferred embodiment of the method of the invention, said (poly)peptides encoded by said recombinant vector molecules of step (b)/(ab) or (c)/(ac) are linked to at least one screenable or selectable tag. In this context, the term screenable or selectable tag refers to a short sequence of amino acids which can be recognized and bound by a particular substance. Tags are commonly used for the purification of biomolecules: examples are His(n), where n = 4-6 which can be bound either by Ni, or a specific antibody, and the flag and myc tags which are recognized by appropriate antibodies. In either of these cases, the tag can be encoded as a C-terminal fusion to all binding partners in library 2. In accordance with the present invention, the tag can be used to isolate e.g. the polyphage referred to

above. Thus, the interaction between the phage-bound binding entity, and its interacting binding partner, establishes a connection between the phage particle and the screenable or selectable tag. This feature can be exploited in a step which relies on e.g. affinity chromatography to isolate the polyphage carrying the interacting molecules. In a final step, those polyphage which carry two distinct nucleic acid molecules and preferably genes (encoding binding entity and binding partner) can be separated from those carrying only one of the two genes e.g. by selection based on transduction or different selectable markers (e.g. antibiotic resistance) present in the individual genomes. In this way, the genes which encode the two interacting molecules can be identified.

A most preferred embodiment of the present invention relates to a method wherein said screenable or selectable tag is encoded by said recombinant vector of step (b)/(ab) or (c)/(ac).

A further most preferred embodiment of the present invention relates to a method wherein said screenable or selectable tag is selected from the list His(n), myc, FLAG, malE, thioredoxin, GST, streptavidin, beta-galactosidase, alkaline phosphatase T7 gene 10, Strep-tag and calmodulin. These screenable tags are all well known in the art and are fully available to the person skilled in the art.

In an additional particularly preferred embodiment of the method of the invention, said screenable or selectable tag is encoded by the genome of the host cell.
An example for this embodiment is an anti-Fc-receptor specific antibody that is expressed by the host cell and could function as an additional bridge in e.g. purification by column chromatography. Another example of this embodiment is an enzyme produced by the host cell that creates a tag such as a phosphorylation on (poly)peptides of the second library without destroying the interaction of (poly)peptides of step (b)/(ab)

with (a)/(aa) so that the modification caused by the enzyme is now the screenable or selectable tag.

In a further preferred embodiment of the method of the invention, said (poly)peptides encoded by the nucleic acid sequences of said additional libraries of step (c)/(ac) cause the interaction of said (poly)peptides of steps (a)/(aa) and (b)/(ab) via phosphorylation, glycosylation, methylation, lipidation or farnesylation of at least one of said (poly)peptides of steps (a)/(aa) and (b)/(ab).

An additional preferred embodiment of the invention relates to a method wherein said host cells in step (d)/(a) are spatially addressable, and the nucleic acid sequences mentioned in step (g)/(d) are retrieved from the corresponding spatially addressable host cell.

In the context of the present invention, the term "spatially addressable" refers to a situation where the individual cells harboring one of the potential combinations of members of the first, second and optionally additional libraries are identifiable by their relative position, e.g. by their position on a master plate. The screening or selection may, for example, be performed either with single clones derived from the master plate, or on a replica plate, thus maintaining the connection between the screenable or selectable property and the information contained in the host cell on the master plate.

An additional preferred embodiment of the invention relates to a method wherein said screenable or selectable property is expressed intracellularly.

Particularly preferred is a method wherein said screenable property is the transactivation of the transcription of a reporter gene such as beta-galactosidase, alkaline phosphatase or nutritional markers such as his3 and leu or resistance genes

giving resistance to an antibiotic such as ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline, or streptomycin.

Furthermore, use can be made of the yeast 2-hybrid system referred to hereinabove or the interaction trap system (Brent et al., EP-A 0 672 131) or of a prokaryotic version analogous to the above recited systems, utilizing the toxR system of Vibrio cholerae (Fritz, H.-J. et al., EP-A 0 630 968). It is within the skills of the person skilled in the art to combine further screening systems known in the art with the method of the present invention.

In a further preferred method of the present invention, said recombinant vectors of step (a)/(aa), (b)/(ab) and (c)/(ac) comprise recombination promoting sites and in said step (e)/(b) recombination events are selected for, wherein said nucleic sequences encoding said (poly)peptides of step (a)/(aa), said nucleic acid sequences encoding said (poly)peptides of step (b)/(ab) and optionally said nucleic acid sequences encoding said (poly)peptides of step (c)/(ac) are contained in the same vector. In this approach, the two genes can be coupled in a single vector, and packaged in a phage of standard size, if appropriate recombination sites are incorporated in the vectors carrying libraries 1 and 2. Again, the phage which carry both nucleic acid sequences and genes are purified with the use of e.g. the screenable tag. If recombination is used to couple the genes from the two libraries, some of the hybrid progeny phage will contain nonrecombinant genomes, since site-specific recombination is not very efficient. However, the hybrid phage can be selected by re-infection of host cells that do not contain library 2 followed by another round of selection of the screenable tag.

In a particularly preferred embodiment of the method of the invention, said recombination events are mediated by the site-specific recombination mechanisms Cre-lox, attP-attB, Mu gin or yeast flp.

In a further particularly preferred embodiment of the method of the invention, said recombination promoting sites are restriction enzyme recognition sites and said recombination event is achieved by cutting the recombinant vector molecules mentioned in steps (a)/(aa), (b)/(ab) and optionally (c)/(ac) with at least two different restriction enzymes and effecting recombination of the nucleic acid sequences contained in said vectors by ligation.

The invention relates in an additional preferred embodiment to a method wherein said identification of said nucleic acid sequences is effected after the selection step (e)/(b) via PCR and preferably sequencing of said nucleic acid sequences after said PCR.
After said selection step (e)/(b), PCR can be carried out with the enriched desired product, conveniently using primers that hybridize to the vector portion of the recombinant vector molecule. Sequencing of the PCR-product may then be carried out according to conventional methods.

In a further preferred embodiment of the method according to the invention, said recombinant vectors of step (a)/(aa), (b)/(ab) and/or (c)/(ac) comprise at least one gene encoding a selection marker.
Said genes encoding said selection markers are preferably different in each of the vectors of step (a)/(aa), (b)/(ab) and/or (c)/(ac), i.e. said vectors comprise genes encoding different selection markers. Said selection markers can conveniently be used for the further purification envisaged in step (f)/(c). For example, a polyphage comprising two members of each library 1 and 2 can be selected for on the basis of a double resistance to antibiotics. Also, a successful recombination event may create a new recombinant vector carrying both nucleic acid molecules from library 1 and 2 as well as genes encoding different selection markers. Again, the selection for a twofold resistance will assist in the identification of the desired product.

In a particularly preferred embodiment of said method, said selection marker is a resistance to an antibiotic, preferably to ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline or streptomycin.

A further preferred embodiment of the present invention relates to a method wherein said host cells are F' and preferably E.coli XL-1 Blue, K91 or its derivatives, TG1, XL1kan or TOP10F.

In a particularly preferred embodiment of the present invention, said RGPs are produced with the use of helper phage taken from the list R408, M13k07 and VCSM13, M13de13, fCA55 and fKN16 or derivatives thereof.

Further preferred is a method wherein at least one of said genetically diverse nucleic acid sequences encode members of the immunoglobulin superfamily.

Said method is particularly preferred, if said genetically diverse nucleic acid sequences encode a repertoire of immunoglobulin heavy or light chains.

In an additional preferred embodiment of the present invention, in said method said genetically diverse nucleic acid sequences are generated by a mutagenesis method. Various mutagenesis methods are well known to the person skilled in the art and need not be described in here in any further detail.

The present invention relates in an additional preferred embodiment to a method in which said genetically diverse nucleic acid sequences are generated from a cDNA library.

In a final preferred embodiment of the method of the invention, said nucleic acid sequences are genes or parts thereof.

As used herein, the term "parts thereof" relates to parts of genes that encode a product that is capable of interacting with a product encoded by any of the other libraries. Thus, it is well known that various proteins are comprised of different domains. Only one of said domains may be capable of interacting with a different (poly)peptide. Such a domain might be encoded by a part of said gene in accordance with the present invention.

The invention also provides for identifying genes encoding more than two interacting peptides or proteins. This can be achieved by using additional vectors encoding genetically diverse additional nucleic acids by an extension of the method described above. As previously, the presence of either a screenable tag or an infectivity protein is used to purify phage carrying genes which encode the components of the complex. Again, the genes in the phage can then be sequenced using methodology well known to those skilled in the art.

Additionally, the present invention relates to a kit comprising at least

(a) a recombinant vector molecule as described in step (a)/(aa) or a corresponding vector molecule;

(b) a recombinant vector molecule as described in step (b)/(ab) or a corresponding vector molecule; and, optionally,

(c) at least one further recombinant vector molecule as described in step (c)/(ac) or a corresponding vector molecule.

As a rule, if recombinant vector molecules are comprised in said kit, they will comprise a library of nucleic acid molecules. In other words, the kit of the invention will contain a plurality of different recombinant vector molecules.

Legends to Figures and Tables

Figure 1: General description of the polyphage principle

a) transform to E. coli hosts

b) infect host containing library1 with helper-phage to package library1 into phage

c) infect cells containing library2 with phages containing library1 leading to cells harboring members of library1 and library2; the presence of library1 and library2 is selected by the presence of the 2 antibiotic resistance markers

d) expression of library1 and library2-tag gene products

e) infect cells with engineered helper-phage to induce polyphage production

Note 1: Polyphage does not discriminate which genome to package therefore the possibilities resulting from step e) arise in an infected cell. To select for the polyphage containing the right packaged genomes the subsequent step is required

f) select for tag e.g., infectivity-mediating protein, in which case ability to infect is selected and

g) select for ability to confer resistance to 2 antibiotics to infected cells

Note 2: Only polyphages that satisfy f) + g) represent phages that display the correct interacting pair and the corresponding genetic information

Figure 2: Co-transformation of two phagemids, polyphage formation and selection *via* His-tag: general description

A, B: libraries of phagemids, preferably with different resistance markers; A: fusions to gIIIp; B: fusions to tag (His); after co-transformation phage production leading to a phage population displaying cognate pairs (left part of the Figure) or not (right part), after selection infection of host cells, selection for double-resistance

Alternative methods include the infection of cells harbouring a plasmid- or phagemid-based library B with a phage library A (prerequisite again: interference-resistant constructs).

Figure 3: pBS vector series: functional map and sequence of pBS13

Figure 4: Co-existence of phagemids: results of restriction digest
Restriction analysis of clones of double resistances (Amp/Cm). R1: pIG10.3, *Xba/ScaI*; R2: pBS13, *Xba/ScaI*, R1+R2: R1 and R2 are mixed in approx. equal proportion; M1: marker λ: *Bst*EII; M2: marker pBR322: *Msp*I; 1 to 10: randomly picked clones: *Xba/ScaI*

Figure 5: Phagemid vector pYING1-C1: functional map
containing the fos peptide. The corresponding vectors pYING1-C2 and pYING1-C3 contain instead of fos the p75 and the IL16 peptides, respectively

Figure 6: Phagemid vector pYANG3-A: functional map
containing the jun peptide. The corresponding vectors pYANG3-Ape2, pYANG3-Ape3, and pYANG3-Ape10 contain instead of jun the p75-binding peptides pe2, pe3, and pe10, respectively

Figure 7: Analysis of selected clones (see Table 2):
7.a: Restriction digest of clones before and after selection
R: pYANG3-Ape2: *Xba*I; M1: marker λ: *Bst*EII; M2: marker pBR322: *Msp*I; α/1 to 10: randomly picked clones before selection: *XbaI/Hin*dIII; β/1 to 10: randomly picked clones after selection: *XbaI/Hin*dIII; size expected: jun-gIII: 745 bp; fos: 256 bp; p75: 577 bp; IL-16: 502 bp
7.b: PCR reaction of clones after selection with primers OPEP5L and OGIII3
R1: pYANG3-A as template; R2: pYANG3-Ape2 as template; M: marker λ: *Bst*EII; β/1 to 10: randomly picked clones after selection as templates

Figure 8: Phagemid vector pING1-C1: functional map
containing the His-tag peptide. The corresponding vector pING3-C1 contains an additional FLAG epitope; pING1-C2 and pING3-C2 contain

the Strep-tag instead of His-tag, with pING3-C2 containing an additional FLAG epitope.

Figure 9: Phagemid vector pONG3-A: functional map for the generation of phage-display libraries (gIII fusions)

Figure 10: Co-transformation of phage and plasmid, polyphage formation and selection *via* SIP: general description

fA: library A in phage construct; B: library B, library members fused to IMP; preferably different resistance markers on phage and plasmid; after co-transformation production of phages; in the case of cognate-pair interaction formation of infectious phages; selection; by plating on double-resistance identification of polyphage particles.

Figure 11: Phage vector fhag1A: functional map for phage-display of the α-HAG scFv

Figure 11a: CAT gene module: functional map and sequence

Figure 12: Phage vector fjun1A: functional map for phage-display of the jun peptide

Figure 13: Phage vector fjun1B: functional map for phage-display of the jun peptide

Figure 14: Phage vector fpep3_1B: functional map for phage-display of the peptide pe3 binding to the intracellular domain of p75

Figure 15: Phage vector fNGF_1B: functional map for phage-display of NGF

Figure 16: Plasmid pUC19/IMPhag: functional map containing fusion of HAG peptide to the N-terminal domains of gIIIp (IMP)

Figure 17: Plasmid pUC18/IMPp75: functional map containing fusion of the intracellular domain of p75 to the N-terminal domains of gIIIp (IMP); pUC18/IMPfos contains the fos peptide instead of the intracellular domain of p75

Figure 18: Plasmid pUC18/IMPIL16: functional map
containing fusion of IL16 to the N-terminal domains of gIIIp (IMP)

Figure 19: Analysis of selected clones (see Table 3)
Lane 1: marker λ: *Bst*EII; lanes 2 to 20: polyphage transductant clones #1 to #19 digested with *Xba/Hind*III; f.._1b: fragment of phage vector after digest; pUC18: fragment of plasmid after digest; α-HAG: fragment containing anti-HAG scFv fused to gIIIc; IMP-p75 and IMP-HAG: fragment containing IMP fused to p75, and IMP-HAG peptide, respectively; pep3-gIIIs: fragment containing pep3 fused to gIIIc (s: short version)

Figure 20: Co-transformation of phagemids, *in vivo* recombination and selection *via* His-tag: general description
A, B: libraries of phagemids; preferably with different resistance markers; A: fusions to gIIIp; B: fusions to tag (His); both constructs containing recombination-promoting sites (*) such as lox/loxP; after co-transformation and recombination production of phages; selection *via* Ni-NTA; re-infection of host cells, selection for double-resistance

Figure 21: *In vitro* recombination and selection *via* His-tag: general description
A, B: libraries of phagemids; preferably with different resistance markers; A: fusions to gIIIp; B: fusions to tag (His); both constructs containing corresponding recognition sites for restriction enzymes (+/o); after digest and co-ligation transformation and production of phages; selection *via* Ni-NTA; re-infection of host cells, selection for double-resistance

Figure 22: Phage vector fjunhag: functional map for phage display of the jun peptide

Figure 23: Spatial *in vivo* SIP: general description
After transformation or co-transformation according to any of the methods described above, a master plate is made. From that phages secreted from individual clones can be analyzed individually (top), or a replica (migration of secreted phages through filter disc) can be made whereon selection for the presence of a tag or infectivity can be performed. By going back to the

master-plate, the information for selected cognate interacting pairs can be retrieved without requiring recombination and/or polyphage production.

Figure 24: *E. coli* display: general description

A, B: libraries of phagemids; preferably with different resistance markers; A: fusions to E.coli surface-display protein; B: fusions to tag (His); after co-transformation expression of constructs; surface-display; in the case of cognate interaction taking place, display of tag on the surface of the host cell; selection

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence

Table 1: Phagemids constructed for Experiments 2 and 3

Table 2: Results of Experiment 2 (see Figure 7)

2.a: Combination of phagemids present in initial library (α)

2.b: Combination of phagemids present after selection (β)

Table 3: Results of Experiment 4 (see Figure 19)

3.a: Identification of phage/plasmid present in individual clones

3.b: Test for infectivity of individual clones

The examples illustrate the invention.

Example 1: General description of the polyphage principle (Figure 1)

The binding entities which comprise library 1 may be peptides or proteins, and are encoded by a genetically diverse collection of first nucleic acid sequences. These nucleic acid sequences are inserted into a first vector which allows for display of the encoded binding entities on the surface of a replicable genetic package. For the purposes of subsequent selection, the first vector should also carry a gene encoding a selectable marker, such as an antibiotic resistance. The binding partners which comprise library 2 may be peptides or proteins, and are encoded by a genetically diverse collection of second nucleic acid sequences which are inserted into a second vector. By way of example, this second vector may be a plasmid, or even a phage or phagemid, in which case the origin of replication should be distinct from that of the first vector. For the purposes of subsequent selection, the second vector should also carry a gene encoding a selectable marker, such as an antibiotic resistance, preferably distinct from that present in the first vector. To facilitate purification of the complex to be formed between any binding entity-binding partner pair, a screenable tag can be conveniently attached to members of library 2.

The two genetically diverse collections of nucleic acids are then introduced into a population of host cells in such a way that encoded libraries 1 and 2 can be expressed. This can be achieved by either (i) co-transformation of the two vectors, or, as actually shown in the figure, (ii) packaging one of the collections of nucleic acids into a vector (such as a bacteriophage) which can be used to infect with high efficiency a population of cells into which the complementary collection of nucleic acid has been introduced. The result is a population of cells in which individual cells carry representatives of each library.

Expression of the two collections of nucleic acids results in the production of pairs of molecules, one from each library, in the host cells. In each case, one or more members

of the library of binding entities is incorporated into the coat of an RGP. In some cells, an interaction will be established between a binding partner on the surface of the RGP and a binding partner expressed from library 2. When such an interaction is established, the RGP therefore carries both the binding entity and the binding partner.

The RGPs displaying such an interaction can then be further purified with the help of polyphage and differing selection markers, as has been discussed hereinabove. After such selection, the nucleic acid sequences encoding one or both binding partners can be conveniently identified by methodology known in the art, such as DNA sequencing.

Example 2: Co-transformation of phagemids with same *E. coli* origin of replication, polyphage formation, and selection of correct pairing interactions *via* His-tag

2.1: Principle (see Figure 2)

To demonstrate that polyphage formation allows the retrieval of the genetic information for cognate protein pairs selected using a tag fused to one member of the protein pair, two separate, small libraries in phagemid vectors are constructed.

2.2: Test of co-existence of phagemids with the same *E. coli* origin of replication:

Prerequisite for the formation of polyphage particles containing two different phagemids is that the different phagemid vectors can co-exist in the host cell.

The vector pBS13 is a derivative of the vector (Krebber *et al.*, 1996) containing a chloramphenicol-resistance gene instead of the kanamycin-resistance gene and a beta-lactamase gene cassette instead of the 2H10-gIII fusion gene, and can be assembled by standard methods starting from pto2H10a3s. Figure 3 contains the functional map and the sequence of pBS13. pIGHAG1A (see Example 4.2.1.f) is digested with *Xba*I and *Hin*dIII. The 1.3 kb fragment containing the anti-HAG gene fused with the C-

terminal domain of filamentous phage pIII protein is isolated and ligated with a pre-digested phagemid vectors pIG10.3, and pBS13 (XbaI-HindIII) to create the vectors pIG10.3-scFv(anti-HAG) (Ap^R) and pBS13-scFv(anti-HAG) (Cm^R), respectively. The vectors are used to transform competent XL-1 Blue cells and selected on LB plates containing Amp/Cm/Tet and glucose (20 mM).

The phagemids from clones of double-resistant colonies (Amp/Cm) are isolated. The restriction digestions indicate the co-isolation of both phagemids from the single colonies (Figure 4).

2.3: Design of libraries A and B:

Library A contains three cyclic peptides each binding to the intracellular domain of the low affinity nerve growth factor (NGF) receptor (see Example 4), and a leucine zipper domain derived from the jun transcription factor, all N-terminally fused to the C-terminal domain of gIII from filamentous phage.

Library B encodes 3 members, namely the leucine zipper domain of the fos transcription factor which heterodimerizes with jun *via* this domain, the intracellular domain of the NGF receptor p75, and, as a negative control which does not interact with library A members, IL-16, all fused at the N-terminus with a His_6-peptide as tag (Hochuli *et al.*, 1988; Lindner *et al.*, 1992).

The cognate pairings are from the interaction between jun and fos (Crameri and Suter, 1993), and p75 and selected cyclic peptides (see Example 4). A non-cognate pairing would occur among the non-cognate pairs mentioned and among jun, or one of the cyclic peptides, and IL-16.

2.4: PCR amplification of the individual constructs

Fos, N-terminus fused to His_6, is PCR amplified using **pOK1** (Gramatikoff *et al.*, 1994) as template and oligonucleotides OFOS-5 and OFOS-3 as primers, where His_6 is

encoded in the OFOS-5 primer. Jun is PCR amplified using pOK1 as template and oligonucleotides OJUN-5 and OJUN-3 as primers.

OFOS-5 5'- GGG*GATATC*CACCACCACCACCACCACCTGCGGTGGTCTGACC

OFOS-3 5'- GGG*GAATTC*CAACCACCGTGTGCCG

OJUN-5 5'- GGG*GATATC*GGTGGTCGGATCGCC

OJUN-3 5'- GGG*GAATTC*ACCACCGTGGTTCATGAC

The hot-start procedure is used. A step-wise touch-down PCR is applied: 92°C, 1 min; 58-52°C, ΔT = 2°C, 1 min; 72°C, 1 min. This is followed by 26 cycles (92°C, 1 min; 52°C, 1 min; 72°C, 1 min).

The PCR products are purified using QIAquick kit (Qiagen) and eluted in ddH_2O. They are then overnight digested with *Eco*RI and *Eco*RV.

The p75 fragment is also PCR amplified using pUC18-IMPp75 (see Example 4) as template and oligonucleotides OP75-5 (where His_6 is encoded) and OP75-3 as primers:

OP75-5 5'- GGG*GATATC*CACCACCACCACCACCACAAGAGGTGGAACAGC

OP75-3 5'- GGG*GAATTC*CACTGGGGATGTGGCAG

The same PCR and restriction digestion conditions as above are applied.

The IL-16 fragment is amplified from the cDNA clone **pcDNA3-ILHu1** (M. Baier, Paul Ehrlich Institute, Germany; Baier *et al.*, 1995; Bannert *et al.*, 1996), using OIL16-5 (where His_6 is encoded) and OIL16-3 as primers.

OIL16-5 5'- GGG*GATATC*CACCACCACCACCACCACCCCGACCTCAACTCCTC

OIL16-3 5'- GGG*GAATTC*GGAGTCTCCAGCAGCTG

The same PCR and restriction digestion conditions as above are applied.

In all cases, the fragments are readily amplified and digested.

2.5: Cloning into intermediate vectors

The digested PCR fragments are gel-purified (QIAquick kit, Qiagen) and eluted into TE buffer. The *Eco*RV/*Eco*RI fragment of pIG1 vector (Ge *et al.*, 1995) is also isolated. The digested PCR fragments of fos, p75, and IL-16 are ligated into the vector fragment, and the ligated vectors transformed into TG1 cells.

The constructs in the pIG1 vector contains the OmpA signal sequence fused in-frame with the constructs.

The correct clones are screened and confirmed by sequencing. They are then *Xba*I/*Hin*dIII digested, and the fragments are isolated.

2.6: Cloning into the expression vectors

The isolated fragments from 2.3 are inserted into pBS13 also excised with *Xba*I/*Hin*dIII, resulting in vectors **pYING1-C1** (Fos), **pYING1-C2** (p75), **pYING1-C3** (IL-16) (see Figure 5). The fragment containing jun is cloned into pIG10.3 vector *via* *Eco*RV/*Eco*RI resulting in **pYANG3-A** (see Figure 6). The anti-p75 peptides pe2, pe3 and pe10 (see Example 4) are cloned into pIG10.3 via *Xba*I/*Hin*dIII, resulting in vectors **pYANG3-Ape2**, **-Ape3** and **-Ape10**, respectively (see Figure 6).

2.7: Selection of correct pairing *via* His-tag

TG1 cells are transformed with the combination of pYANG3-A + pYING1-C1, or pYANG3-A + pYING1-C2, or pYANG3-A + pYING1-C3, or (pYANG3-Ape2, -Ape3 and -Ape10) + pYING1-C1, or (pYANG3-Ape2, -Ape3 and -Ape10) + pYING1-C2, or (pYANG3-Ape2, -Ape3 and -Ape10) + pYING1-C3, thus creating all possible combinations separately to ensure the presence of each of them in the selection experiment. The transformed cells are plated on ampicillin/chloramphenicol-containing LB agar plates, and colonies with double resistance (Ap^R/Cm^R) are selected.

The colonies are scraped off the plates and used to inoculate 2xYT medium (Amp/Cm) and shaken at 37°C for 3 hrs. The cultures are induced (1 mM IPTG) at 30°C for 1 hr and infected with R408 (Stratagene) at 37°C for 30 min. The cultures are shaken at RT for 3 hrs, kanamycin is added and shaking continued at RT overnight.

The phage particles are harvested from the overnight cultures, mixed and PEG-precipitated. The phages are directly selected on immobilized Ni-NTA (NI-NTA HisSorb Strips, Qiagen). The eluted phages are used to infect TG1 cells, which are plated on ampicillin/chloramphenicol-containing LB agar plates, and colonies with double resistance (Ap^R/Cm^R) are selected.

The phagemids of selected clones are isolated and analyzed by restriction digest (see Figure 7.a) and used as templates for PCR screening. Primer OPEP5L is used to amplify the pYANG3-Ape2, -Ape3 and -Ape10 constructs specifically (see Figure 7.b).

OPEP5L 5'- GACTACAAAGATGTCGACTG

There is a specific enrichment of constructs of correct pairing (Table 2).

Example 3: Interactive screening of *E. coli* genomic DNA libraries (Polyphage/tag system)

3.1: Principle (see Figure 2)

Instead of using two model libraries as in Example 2, a genomic DNA library of *E. coli* is prepared to be screened against itself to identify interacting *E. coli* peptides or proteins.

3.2: Construction of display and expression vectors for genomic DNA

Expression vectors are constructed having a blunt-end restriction site SmaI inserted either in front of His-tag, Strep-tag (Schmidt and Skerra, 1994) or the C-terminal domain of gIII (gIIIc) *via* oligonucleotide cassettes or PCR.

The self-complementary oligonucleotides OHIS5 & OHIS3, and OSTREP5 & OSTREP3, are used to create ds DNA cassettes encoding the His-tag, and the Strep-tag, respectively.

OHIS5 5'- AATTCCCCGGGCACCACCACCACCACCACTGATA

OHIS3 5'- AGCTTATCAGTGGTGGTGGTGGTGGTGCCCGGGG

OSTREP5 5'- AATTCCCCGGGTCTGCTTGGCGTCACCCGCAGTTCGGTGGT-TGATA

OSTREP3 5'- AGCTTATCAACCACCGAACTGCGGGTGACGCCAAGCAGACC-CGGGG

The cassettes upon phosphorylation and annealing recreate the *Eco*RI and *Hin*dIII sites. The cassettes are inserted into pIG1 and pIG3 vectors (Ge *et al.*, 1995) cut by the same restriction enzymes. The resulting vectors are pING1-A1, pING3-A1 (for His tag in pIG1 and pIG3 vectors) and pING1-A2, pING3-A2 (for Strep-tag), respectively. The correct vectors are screened for the presence of *Xma*I site (isoschizomer of *Sma*I) and the constructs are confirmed by sequencing. The *Xba*I/*Hin*dIII fragments of these vectors are inserted into pBS13 vector, linearized with the same enzymes, resulting in vectors **pING1-C1**, **pING3-C1** and **pING1-C2**, **pING3-C2,** respectively (see Figure 8).

The gIIIc fragment containing the SmaI site is generated from PCR amplification of pIG10.3 vector using primers OGIII5 and OGIII3, where OGIII3 anneals 3' of the gene III in the vector:

OGIII5 5'- CG*GAATTC*CCCGGGGAGCAGAAGCTGATC

OGIII3 5'- TTTTTCACTTCACAGGTC

Three rounds of PCR are performed with a hot-start: 92°C, 1 min; 46°C, 1 min; 72°C, 1.5 min. This is followed by 30 rounds of: 92°C, 1 min; 50°C, 1 min; 72°C, 1.5 min.

The PCR product is purified (QIAquick) and digested with *Eco*RI and *Hin*dIII. The fragment is gel-purified (QIAquick) and ligated into pIG10.3. The sequence of the resulting vector, **pONG3-A** (see Figure 8), is confirmed by restriction analysis and by sequencing.

3.2: Selection of Interacting Pairs from *E. coli* Genomic DNA via His-tag

Genomic DNA of *E. coli* strain XL-1 Blue (Stratagene) is isolated using the Blood & Cell Culture DNA Maxi kit (Qiagen) and eluted in TE buffer (pH 8.0). 200 μg of the DNA is

taken and sonicated (50 cycles, 270 mA, 0.5 s/stroke). The fragmented DNA (average size: max. 0.7 kB) is blunt-ended by a fill-in reaction with T4 DNA polymerase.

Vectors pING1-C1 and pONG3-A are digested with *Eco*RV and *Sma*I, the vector fragments are gel-purified (Qiagen). The vector fragments are then ligated with the blunt-ended genomic DNA at 16°C overnight. The ligation mixtures are taken to transform TG1 cells.

The pING1-C1 and pONG3-A transformants are scratched from the plate and used to inoculate 2xYT medium containing Cm/glucose or Amp/glucose, respectively. The pING1-C1 culture is infected with helper-phage (VCSM13 or M13k07) and phage particles are isolated. These phage particles are used to infect log-phase cells containing the pONG3-A library. The resulting culture is plated out on large Amp/Cm/glucose plates.

The colonies are scratched from the surface of the plates above and transferred to 2xYT medium containing Amp/Cm. After 30 min shaking at 37°C, the culture is then induced (1 mM IPTG) for 30 min, infected with helper-phage at 37°C for 30 min and shaken at RT overnight.

The phage particles are harvested from the overnight culture and PEG-precipitated. They are selected on immobilized Ni-NTA (NI-NTA HisSorb Strips, Qiagen). The eluted phages are used to infect log-phase TG1 cells. Selected protein pairs are characterised by determination of their corresponding DNA sequences.

Example 4: Polyphages and Selection of Correct Pairing Interactions *via* SIP

4.1: Principle (see Figure 10)

The purpose of this experiment is to show that from a combination of 2 libraries one can isolate and identify the correct interacting pairs using the SIP (Selectively Infective Phage: Krebber *et al.*, 1995; the term "IMP" used in the experimental section denotes "Infectivity mediating particle" comprising the N-terminal domains of the gene III protein

of filamentous phage) selection system, and recover the information about both interacting partners *via* the formation and selection of polyphage particles. The library members forming interacting pairs with members of the corresponding library are being 'doped' with library members that do not interact with members of the corresponding library, and thus should not give a positive SIP selection.

4.2: Construction of vectors

4.2.1: fhag1A (see Figure 11)

a. The phage vector f17/9-hag (Krebber *et al.*, 1995) is digested with EcoRV and XmnI. The 1.1 kb fragment containing the anti-HAG Ab gene is isolated by agarose gel electrophoresis and purified with a Qiagen gel extraction kit. This fragment is ligated into a pre-digested pIG10.3 vector (EcoRV-XmnI). Ligated DNA is transformed into DH5a cells and positive clones are verified by restriction analysis. The recombinant clone is called **pIGhag1A**. All cloning described above and subsequently are according to standard protocols (Sambrook *et al.*, 1989)

b. The vector f17/9-hag (Krebber *et al.*, 1995) is digested with EcoRV and StuI. The 7.9 kb fragment is isolated and self-ligated to form the vector **fhag2**.

c. The chloramphenicol resistance gene (CAT) assembled *via* assembly PCR (Ge and Rudolph, 1997) using the the template pACYC (Cardoso and Schwarz, 1992) (Figure 11a shows the functional map and the sequence of the CAT gene) is amplified by the polymerase chain reaction (PCR) with the primers:

CAT_BspEI(for): 5' GAATGCTCATCCGGAGTTC

CAT_Bsu36I(rev): 5' TTTCACTGGCCTCAGGCTAGCACCAGGCGTTTAAG

d. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with BspEI and Bsu36I then ligated into pre-digested fhag2 vector (BspEI-Bsu36I; 7.2 kb fragment) to form **fhag2C**.

e. The vector fhag2C is digested with EcoRI and the ends made blunt by filling-in with Klenow fragment. The flushed vector is self-ligated to form vector **fhag2CdelEcoRI**.

f. pIGHAG1A is digested with XbaI and HindIII. The 1.3 kb fragment containing the anti-HAG gene fused with the C-terminal domain of filamentous phage pIII protein is isolated and ligated with a pre-digested fhag2CdelEcoRI phage vector (XbaI-HindIII; 6.4 kb) to create the vector **fhag1A**

4.2.2: fjun1A (see Figure 12)

a. The EcoRV site of pIG10.3 is converted to a SalI site by oligonucleotide site-directed mutagenesis (Sambrook *et al.*, 1989) with primer:
SalI9-9primer(rev) 5'CTGAATGTCGACATCTTTGTAGTC3'
The mutated pIG10.3 is called **pIG10.3 SalI.**

b. The jun leucine-zipper domain from **pOK1** (Grammatikoff *et al.*, 1994) is amplified by PCR with the primers:
jun2(for): 5'ACGCGTCGACGCCGGTGGTCGGATCGCCCGG3'
jun2(rev): 5'AATTCGGCACCACCGTGGTTCATGACT3'

c. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with SalI and EcoRI, then ligated into pre-digested pIG10.3SalI vector (SalI-EcoRI) to form the vector **jun-pIG10.3SalI.**

d. The vector jun-pIG10.3SalI is digested with XbaI and EcoRI. The 0.14 kb fragment is ligated into the pre-digested vector fhag1A (XbaI-EcoRI; 7kb) to form the vector **fjun1A**.

4.2.3: fjun1B (see Figure 13)

a. The DNA encoding the C-terminal domain including the long linker separating it from the amino terminal domain of the filamentous phage pIII (gIII short) is amplified by PCR using pOK1 (Grammatikoff *et al.*, 1994) as template with the primers:
gIII short(for): 5'GCTTCCGGAGAATTCAATGCTGGCGGCGGCTCT3'
gIII short(rev): 5'CCCCCCCAAGCTTATCAAGACTCCTTATTACG3'

b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with EcoRI and HindIII, then ligated into pre-digested fhag1A vector (EcoRI-HindIII) to form the vector **fjun1B.**

4.2.4: fpep2_1b, fpep3_1B, fpep10_1b (see Figure 14)

a. These constructs are obtained from a peptide library screened against the intracellular domain of p75, the low affinity receptor of NGF, in a SIP experiment.

b. A peptide library cassette of cyclic peptides with length variants of 6-16 amino acids is prepared from the oligos:

Groprim: 5'-CATGAATTCGGATCCTCC-3'

Gron10: 5'-CTATGGCGCGCCTGTCGACTGT$(M)_{6\text{-}16}$TGTGGTGGTGGAGGATC-CGAATTCATG-3'

where M is a mixture of 19 trinucleotide codons (Virnekäs *et al.*, 1994), excluding the one coding for Cys. The length variation is achieved by coupling 6 trinucleotide positions using the standard coupling procedure, and, for the next 10 coupling cycles, by omitting the capping step during DNA synthesis and by diluting the trinucleotide mixture to achieve stepwise coupling yields of 50%.

The oligos are annealed and filled in with the Klenow fragment of DNA polymerase I to form a double-stranded DNA cassette with standard methods (Sambrook *et al.*, 1989). The cassette is digested with SalI-EcoRI, purified with Qiaex DNA gel extraction kit, and ligated to pre-digested fjun1B vector (SalI-EcoRI) to form the peptide library. The ligated peptide library is transformed into competent DH5a cells harboring pUC18/IMP-p75 (see below) and plated on Luria Broth (LB) (30 μg/ml chloramphenicol + 100 μg/ml ampicillin) and incubated overnight at ambient temperature.

c. The Amp^r Cm^r colonies are scraped with LB, and 1 ml of suspension is used to inoculate 25 ml LB (30 μg/ml chloramphenicol + 100 μg/ml ampicillin + 1 mM IPTG). The culture is incubated overnight at room temperature.

d. The supernatant is separated from the cells by centrifugation (10,000 RPM, 10 min., 4°C). 5 ml of 30% PEG/3M NaCl are added to the supernatant and mixed 100 times. After 1 hour on ice, the phage precipitate is collected by centrifugation (10,000 RPM, 10 min., 4C). The pellet is resuspended in 1 ml TBS buffer. The suspension is filtered with a 0.45 micron filter (Sartorius).

e. 100 µl of log phase K91 cells (or any male E. coli cells (F-pilus containing)) are infected with 10 µl of phage supernatant, plated on LB (30 µg/ml chloramphenicol) and incubated overnight at ambient temperature.

f. Chloramphenicol-resistant transductants are picked, and overnight cultures are prepared to isolate DNA for sequencing. From the sequencing, fpep2_1b, fpep3_1B, fpep10_1b containing peptides pe2, pe3, and pe10 are identified.

pe2: 5'-TGTTTTTTTCGTGGTGGTTTTTTTAATCATAATCCTCGTTATTGT-3'

(CysPhePheArgGlyGlyPhePheAsnHisAsnProArgTyrCys)

pe3: 5'-TGTATTGTTTATCATGCTCATTATCTTGTTGCTAAGTGT-3'

(CysIleValTyrHisAlaHisTyrLeuValAlaLysCys)

pe10: 5'-TGTTCTTATCATCGTCTTTCTACTCGTGTTTGT-3'

(CysSerTyrHisArgLeuSerThrArgValCys)

4.2.5: fNGF1B (see Figure 15)

a. The DNA encoding the nerve growth factor (NGFI) gene is amplified from pXM NGF (Ibanez *et al.*, 1992) as template with the primers:

NGF(for): 5'AAAAAAGTCGACTCATCCACCCACCCAGTC3'

NGF(rev): 5'AGGAATTCGCCTCTTCTTGCAGCCTT3'

b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with SalI and EcoRI, then ligated into pre-digested fjun1B vector (SalI-EcoRI) to form the vector **fNGF1B.**

4.2.6: pUC19/IMP-HAG (see Figure 16)

a. The vector f17/9-hag (Krebber *et al.*, 1995) is digested with EcoRI and HindIII. The 1.4 kb fragment containing the gene fusion of the IMP with the HAG peptide, is isolated and cloned into pre-digested **pUC19** (EcoRI-HindIII) to form the vector **pUC19/IMP-HAG**

4.2.7: pUC18/IMP-p75 (see Figure 17)

a. The intracellular domain of p75 containing the C-terminal 142 amino acids is amplified from the cDNA clone of p75 (Chao *et al.*, 1986) as template with the primers:

p75(for): 5' GCTGGCCCGTACGACAAGAGGTGGAACAGCTGC

p75(rev): 5' TCTCGAAGCTTATCACACTGGGGATGTGGC

b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with BsiWI and HindIII, then ligated into pre-digested pUC19 vector (BsiWI-HindIII) to form the vector **pUC19/IMP-p75.**

c. The vector pUC19/IMP-p75 is digested with XbaI and HindIII. The 1 kb fragment is isolated and cloned into the pre-digested **pUC18** vector (XbaI-HindIII) to form the vector **pUC18/IMP-p75.**

4.2.8: pUC18/IMP-IL16 (see Figure 18)

a. The IL16 gene is amplified from the clone **pcDNA3-ILHu1** (M. Baier, Paul Ehrlich Institute, Germany; Baier *et al.*, 1995; Bannert *et al.*, 1996) as template with the primers:

f1Bsu36Ifor: 5'AGACTGCCTCAGGCCAGCCCGACCTCAACTCC3'

f3HindIIIrev2: 5'ATATATAAGCTTTTAGGAGTCTCCAGCAGC3'

b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with Bsu36I and HindIII, then ligated into pre-digested pUC18/IMP-p75 vector (Bsu36I-HindIII) to form the vector **pUC18/IMP-IL16.**

4.3: In vivo SIP with co-transformation and polyphage

4.3.1: Combining 2 libraries (Library 1 is fused with gIII while Library 2 is fused to the IMP).

10 ng each of fjun1B, fjun1A, fpep3_1B, fhag1A, fNGF1B with 500 ng each of pUC18/IMP-p75, pUC18/IMP-HAG, pUC18/IMP-IL16 are co-transformed into DH5a cells by electroporation. The cells are plated on Luria Broth (LB) (30 μg/ml chloramphenicol + 100 μg/ml ampicillin) and incubated overnight at ambient temperature.

The Amp^r Cm^r colonies are scraped with LB and 1 ml of suspension is used to inoculate 25 ml LB (30 μg/ml chloramphenicol + 100 μg/ml ampicillin + 1 mM IPTG) followed by incubation overnight at room temperature.

4.3.2: In vivo SIP. The supernatant from the cells is separated by centrifugation (10,000 RPM, 10 min., 4°C). 5 ml of 30% PEG/3M NaCl are added to the supernatant and mixed 100 times. After 1 hour on ice, the phage precipitate is collected by centrifugation (10,000 RPM, 10 min., 4°C). The pellet is resuspended in 1 ml TBS buffer, and the suspension is filtered through a 0.45 micron filter (Sartorius).

200 μl of phage supernatant are used to infect 1.8ml of log phase K91 cells (or any male E. coli cells (F-pilus containing)), and the cells are plated on LB (30 μg/ml chloramphenicol + 100 μg/ml ampicillin) and incubated overnight at ambient temperature.

4.3.3: Testing of infectious polyphage DNA patterns and infectity. Twenty individual Amp^r Cm^r colonies are used to inoculate 5 ml LB (30 μg/ml chloramphenicol + 100 μg/ml ampicillin) in each case and incubated at ambient temperature overnight. Plasmid and RF DNA are isolated from each clone with a Qiagen Miniprep DNA kit. Clones are analysed by restriction analysis with restriction enzymes XbaI and HindIII together with appropriate buffers as supplied and instructed by the manufacturer. The restriction

digests are run in a 0.8% TBE agarose gel at constant voltage of 100V for 1.5 hours. The restriction patterns, together with the relative intensity of the bands (because the phage vectors (fjun1B, fjun1A, fpep3_1B, fNGF1B, fhag1A) have significantly lower copy numbers than the plasmid vectors) allow to identify correctly interacting pairs. For the pair fhag1A+pUC19/IMP-HAG, an XbaI-HindIII digest will yield a 6.5 kb, 3.3 kb, 1.3 kb, and 0.7 kb fragments, while for the pair fpep3_1B+pUC18/IMP-p75, the same digest will yield 6.3 kb, 2.8 kb, 1kb, and 0.7kb fragments. A problem though is to distinguish the potential non-cognate combinations of fjun1B or fjun1A with pUC18/IMP-p75 because they would give similar patterns as the fpep3_1B+pUC18/IMP-p75. To further resolve this, the clones containing identical patterns can be re-digested with BamHI-HindIII. The fjun1A or fjun1B in combination with pUC18/IMP-p75 would yield only 4 fragments - 4.1 kb and 2.9 kb , 2.6 kb , 1.2 kb fragments - while the cognate pair fpep3_1B+pUC18/IMP-p75 will yield 5 fragments - 3.5 kb, 2.9 kb, 2.6 kb, 1.2 kb, 0.5 kb.

To further prove that cognate interacting pairs have been selected, the ability of the clones to form selectively-infective phage particles is tested. Only clones with a cognate pair can form infectious phages. The supernatant from the overnight culture of the individual clones is filtered with a 0.45 micron filter (Sartorius). Ten microliters of phage supernatant are mixed with 100 µl of log phase K91 cells (or any male E. coli cells (F-pilus containing)) for 10 minutes at 37°C. The suspension is plated on LB (30 µg/ml chloramphenicol) and incubated overnight at 37°C. The result is shown in Table 3.b.

In summary (see Figure 19), the results from the above example indicate that among 19 clones analyzed, 8/19 have the cognate pair fpep3_1B+pUC18/IMP-p75 and produce selectively-infective phage; 1/19 has the fhag1A+pUC19/IMP-HAG combination and produces selectively-infective phage.

Example 5: Combination of Multiple Libraries into a Single Phagemid Vector through Recombination, Screening *via* tag system

5.1: Principle (see Figure 20)

To be able to retrieve the genetic information for cognate protein pairs selected *via* a tag fused to one of the partners, two separate libraries in phagemid vectors are constructed containing the *lox* recombination promoting sites and recombined on one phagemid by action of the *cre* recombinase in an *in vivo* recombination.

5.2: Vector construction

Both *lox*P and *lox*P511sites (Hoess *et al.*, 1986) are inserted in tandem into the region flanked by the ColE1 ori and β-lactamase in vector pING1-C1, whereas in vector pONG3-A, the *lox*P site is cloned upstream of the *Xba*I site and the *lox*P511 downstream of the *Hind*III site. Therefore, the genomic DNAs to be cloned are flanked by the *lox*P and *lox*P511 sites.

5.3: Library construction and recombination

The libraries are prepared as in Example 3. The phagemids in the double-resistant clones are recombined through the *cre* recombinase which either is encoded in the phagemid being inducible (Tsurushita *et al.*, 1996), or is transferred through P1 phage infection (Rosner, 1972; Waterhouse *et al.*, 1993). Phages are prepared from the recombined clones by helper phage infection and used to infect new *E. coli* cells (*cre*$^-$).

5.4: Selection

The phage particles are prepared from the Cm^R clones and subjected to His-tag selection as in Examples 2 and 3. The sequences encoded in each phagemid, which now contains members of both libraries, can be determined by sequencing using primers specific for myc-tag region (library 1) and His-tag region (library 2).

Example 6: SIP-based library vs. library screening via *in vitro* recombination of separately constructed libraries into one phage vector

6.1: Principle (see Figure 21)

To be able to retrieve the genetic information for cognate protein pairs selected by SIP interaction *in vivo*, two separate libraries in phage and plasmid vectors are constructed and recombined by co-ligation in an *in vitro* recombination.

6.2: Construction of Libraries A and B

Library A encodes 2 members, namely a single chain Fv antibody against a peptide derived from hemagglutinin (fαhag) and the leucine zipper domain derived from the jun transcription factor (fjun), both N-terminally fused to the C-terminal domain of gIII from filamentous phage and preceded by the ompA signal sequence followed by the Flag epitope.

Library B encodes 3 members on plasmid vectors of the pUC series, namely the hemagglutinin peptide to which the above αhag antibody binds (pUC19-IMPhag), the leucine zipper domain of the fos transcription factor (pUC18-IMPfos) which heterodimerizes with jun via this domain, and the intracellular domain of the low affinity nerve growth factor receptor (pUC18-IMPp75), as a negative control which does not interact with library A members, all fused to the infectivity-mediating N-terminal domains of phage gIII protein, preceded by the gIII signal sequence.

Library A members are cloned into a fd phage vector which also contains downstream of the library A insertion site the N-terminal domains (N1-N2) of gIII, followed by the cloning sites *Bsi*WI and *Hin*dIII to allow in-frame insertion of library B members.
Library A construct fαhag is identical to the f17/9-hag fd phage vector (Krebber *et al.*, 1995) and serves as basis for construction of fjun. The jun leucine zipper together with amino acids 290 to 326 of the C-terminal part of gIII is PCR-amplified (primers FR620 and FR621, containing *Eco*RV and *Sfi*I sites, respectively) from the construct fjun1B (containing the jun leucine zipper fused to amino acids 290 to 493 of gIII) generated in Example 4. The resulting PCR fragment is ligated directionally into EcoRV/SfiI-digested

f17/9-hag vector in frame with amino acids 327 to 493 of the gIII C-terminal domain resulting in vector **fjunhag** (see Figure 22).

Generation of library B constructs pUC19-IMPhag and pUC18-IMPp75 is described in Example 4. To construct pUC18-IMPfos, amino acids 219 to 272 of the N-terminal part of gIII together with the fos leucine zipper are PCR-amplified (primers FR618 and FR619, containing BsiWI and HindIII sites, respectively) from the pOK1 phagemid vector (Grammatikoff *et al.*, 1994). The resulting PCR fragment is ligated directionally into BsiWI/HindIII-digested pUC18-IMPp75 to create **pUC18-IMPfos** (see Figure 17).

Primers:

FR618: 5'CGCCGTACGGCGGCTCTGGTGGTGGTTCTGGTGGC3'

FR619: 5'CCCAAGCTTTTAGACTAGCTGACTAGAAGATCTGC3'

FR620: 5'CGCGATATCGTCGACGCCGGTGGTCGGATCGCC3'

FR621: 5'CGCGGCCCCCGAGGCCCCACCACCGGAACCGCCTCCC3'

6.3: Preparation and recombination of library A and B and selection of interacting protein pairs by SIP

Non-covalent, cognate interactions of αhag antibody with hag peptide (Krebber *et al.* 1995) and of fos and jun leucine zipper domains (Grammatikoff *et al.*, 1994) generates infective SIP phage. Thus, from the six possible combinations of members of the model libraries A and B (**fαhag-hag**, fαhag-fos, fαhag-p75, **fjun-fos**, fjun-hag, fjun-p75), only two combinations (cognate pairs in bold) should be selected by *in vivo* SIP. To recombine the library members in all possible permutations, library A is linearized by digestion with BsiWI/HindIII to prepare it for random incorporation of library B members, prepared by mass-excision with BsiWI/HindIII from the construct B pool described above. After co-ligation of the mass-excised library B fragments into library A vectors, the sample is transformed into competent E.coli cells, plated onto chloramphenicol-containing LB agar plates and grown overnight at 37°C. The recombined library size can be determined by plating serial dilutions of the transformation and can be compared to

the complexities of the individual libraries A and B. The total recombined library is scraped from the plates in LB medium and used to inoculate an appropriate volume of chloramphenicol-selective LB-medium supplemented with 1 mM IPTG. After growth at 30°C overnight with constant shaking to allow production of SIP phages, the bacteria are pelleted by centrifugation and phages present in the supernatant are precipitated on ice for one hour by addition of 0.25 volumes of 20% PEG/2.5 M NaCl. The phages are pelleted by centrifugation for 30 min at 10 000 x g and 4°C. The pellet is resuspended in an appropriate volume of 1 x TBS buffer and filtered through a 0.45 µM filter. Serial dilutions of this filtrate are used to infect F^+ E.coli cells. The double-stranded, replicative form phage DNA is prepared from resulting transductant colonies by standard methods and analyzed by restriction digest and sequencing for the presence and identity of library A and B members. Furthermore, the supernatant of transductant colonies is analyzed for the presence of infective SIP phages to confirm that protein-protein interaction of a particular pair selected from the recombined libraries A and B is responsible for SIP phage infectivity.

Alternatively, the model libraries A (2 members) and B (3 members) are used to construct all possible combinations (listed above) individually, and equal amounts (50 ng) of each of the 6 combinations can be co-transformed into competent E. coli cells followed by the steps listed above. The distribution of individual constructs after co-transformation as well as the distribution of transductants resulting from the model library can be analyzed as described above. The selective recovery of phage constructs which co-encode cognate protein pairs demonstrates the feasibility of SIP-based selection of binding partners after an appropriate recombination event.

Example 7: 'Spatial' *in vivo* SIP

7.1: Principle (see Figure 23)

Coupling of information about members of interacting peptides or proteins is achieved by having a spatial relationship between the particles displaying the selectable or

screenable property (in this example phages for the SIP experiment) and the package containing the genetic information for the individual library members (in this example the *E. coli* cell secreting the phage particle being screened), i. e. a correlation between the phage being examined and the position of the corresponding *E. coli* host on the master plate.

7.2: Combining 2 libraries (Library A is fused with gIII while library B is fused to the IMP)

10 ng each of fjun1B, fjun1A, fpep3_1B, fhag1A, fNGF1B are co-transformed with 500 ng each of pUC18/IMP-p75, pUC19/IMP-HAG, pUC18/IMP-IL16 into DH5a cells by electroporation. The transformants are plated on LB (30 μg/ml chloramphenicol + 100 μg/ml ampicillin) and incubated overnight at ambient temperature.

7.3: Screening of co-transformants by SIP

From the master plate of co-transformants, each of the co-transformants are labelled and inoculated separately into 5 ml LB (30 μg/ml chloramphenicol + 100 μg/ml ampicillin) and incubated overnight at ambient temperature.

Plasmid and RF DNA are isolated from each clones with a Qiagen Miniprep DNA kit. Clones are analysed by restriction analysis with restriction enzymes XbaI and HindIII together with appropriate buffers as supplied and instructed by the manufacturer. The restriction digests are run in a 0.8% TBE agarose gel at constant voltage of 100 V for 1 to 2 hours. Restriction patterns allow discrimination of the particular clones.

The supernatant from the overnight culture of the individual clones is filtered with a 0.45 micron filter (Sartorius). Ten microliters of phage supernatant are mixed with 100 μl of log phase K91 cells (or any male E. coli cells (F-pilus containing)) for 10 minutes at 37°C. The suspension is plated on LB (30ug/ml chloramphenicol) and incubated overnight at 37°C.

A positive co-transformant (i.e. contains the correct interacting pair) has a corresponding correct restriction pattern and is capable of producing infectious phages, that are incapable of secondary or subsequent infections. Polyphage particles being capable of such infections, and containing the genetic information of an interacting pair as well, can readily be identified by their restriction digest pattern.

Example 8: *E. coli* display

8.1: Principle (see Figure 24)

Two libraries are introduced into *E.coli* cells, with expressed members of library A (such as antibody, peptide, or cDNA libraries) being presented at the surface of the cells. In those cases where interacting pairs are formed, members of library B (such as antibody, peptide, or cDNA libraries) are transported in the complex with its cognate partner to the surface of the cell as well, thus displaying a selectable or screenable property such as a tag. Selected cells contain the information for both interacting partners.

8.2: Preparation of Library A

A thioredoxin peptide library is prepared as fusions to the *E. coli* flagellin in the pFLITRX vector essentially as described (Lu *et al.*, 1995).

8.3: Preparation of Library B

An cyclic, variable-length peptide library including a FLAG epitope (Hopp *et al.*, 1988; Knappik and Plückthun, 1994) is prepared essentially as described in Example 4.2.4, and cloned in the pTERM vector, a modified version of the pto2H10a3s vector (Krebber *et al.*, 1996) containing a chloramphenicol-resistance gene instead of the kanamycin-resistance gene. The pTERM vector can be assembled by standard methods starting from pto2H10a3s. This cyclic peptide library is packaged by infection with a helper phage (M13K07 or VCSM13) by standard methods (Sambrook *et al.*, 1989).

8.4: Combination of Library A and Library B

An aliquot of the *E. coli* cells containing Library A is used to inoculate 50 ml LB (100 µg/ml ampicillin) and incubated at ambient temperature until the OD600 reached 0.4. The cells are infected with phages containing Library B at a multiplicity of infection (MOI) of 10. After 30 min of infection, the cells are collected by centrifugation (5000 RPM, 10 minutes, 4°C) and resuspended in 1 ml LB. The suspension is plated on M9 media (+ 1 mM $MgCl_2$, supplemented with 0.5% glucose, 0.2% casamino acids, 100 µg/ml ampicillin, 30 µg/ml chloramphenicol).

8.5: Selection of interacting pairs

The Amp^r Cm^r colonies are scraped with M9 media (+ 1 mM $MgCl_2$, supplemented with 0.5% glucose, 0.2% casamino acids, 100 µg/ml ampicillin, 30 µg/ml chloramphenicol), and an aliquot of the suspension is used to inoculate 25 ml M9 media (+ 1 mM $MgCl_2$, supplemented with 0.5% glucose, 0.2% casamino acids, 100 µg/ml ampicillin, 30 µg/ml chloramphenicol) and incubated at 37°C until saturation. Selection is performed essentially as described (Lu *et al.*, 1995), the modification being that the antibody used for selection is the M1 anti-FLAG antibody (Kodak).

Individual enriched Amp^r Cm^r colonies are isolated and the sequences of the corresponding interacting peptide(s) and cyclic peptide(s) are determined by DNA sequencing. To confirm that the encoded peptide and cyclic peptide form a cognate pair, each of the clones is tested for enrichment based on the selection method described above, whereby the Amp^r Cm^r colonies bind to the M1 anti-FLAG antibody in a single round of selection.

Literature:

Baier *et al.*, 1995, *Nature* 378, 563

Bannert *et al.*, 1996, *Nature* 381, 30

Cardoso and Schwarz, *J. Appl. Bacteriol.* 72 (1992) 289-293

Crameri and Suter, 1993, *Gene* 137, 69-75

Ge *et al.*, 1995, *Antibody Engineering*, 2nd ed., 229-266

Ge and Rudolph, *BioTechniques* 22 (1997) 28-29

Gramatikoff *et al.*, *Nucleic Acids Res.* 22 (1994) 5761-5762

Hochuli *et al.*, 1988, *Bio/Technology* 6, 1321-1325

Hoess *et al.* 1986, *Nucleic Acids Res.* 14, 2287-2300

Hopp *et al.*, 1988, *Bio/Technology* 6, 1204-1210

Ibanez *et al.*, 1992, *Cell* 69, 329-341

Knappik and Plückthun, 1994, *BioTechniques* 17, 754-761

Krebber *et al.*, 1995, *FEBS Letters* 377, 227-231

Krebber *et al.*, 1996, *Gene* 178, 71-74

Lindner *et al.*, 1992, *Methods: A companion to Methods Enzymol.* 4, 41-56

Lu *et al.*, 1995, *Bio/Technology* 13, 366-372

Rosner, 1972, *Virology* 48, 679-689

Sambrook *et al.*, 1989, *Molecular Cloning: a Laboratory Manual*, 2nd ed.

Schmidt and Skerra, 1994, *J. Chromatogr. A* 676, 337-345

Tsurushita *et al.*, 1996, *Gene* 172, 59-63

Virnekäs *et al.*, 1994, *Nucleic Acids Res.* 22, 5600-5607

Waterhouse *et al.*, 1993, *Nucleic Acids Res.* 21, 2265-2266

CLAIMS

1. A method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:

   (a) providing a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;

   (b) providing a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as mentioned in step (a), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (a) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (a) and (b), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;

   (c) optionally, providing additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (a) and/or step (b), wherein the vector molecules employed for the

production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (a) and (b) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (c) together with at least one of said properties displayed by either said vector molecule and/or said recombinant inserts used in steps (a) and/or (b), upon the interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;

(d) expressing members of said libraries of recombinant vectors or nucleic acid sequences mentioned in steps (a), (b) and optionally (c), in appropriate host cells so that at least one interaction is established;

(e) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;

(f) optionally, carrying out further screening, selection and/or purification steps; and

(g) identifying said nucleic acid sequences encoding said (poly)peptides.

2. A method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:

(a) expressing in appropriate host cells

(aa) nucleic acid sequences contained in a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;

(ab) nucleic acid sequences contained in a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as mentioned in step (aa), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (aa) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (aa) and (ab), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;

(ac) optionally, nucleic acid sequences contained in additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (aa) and/or step (ab), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (aa) and (ab) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (ac) together with at least one of said properties displayed by either said vector molecule and/or said recombinant inserts used in steps (aa) and/or (ab), upon the

interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;

so that at least one interaction is established;

(b) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;
(c) optionally, carrying out further screening, selection and/or purification steps; and
(d) identifying said nucleic acid sequences encoding said (poly)peptides.

3. The method according to claim 1 or 2, wherein said screenable or selectable property is expressed extracellularly.

4. The method according to any one of claims 1 to 3 wherein said recombinant vector molecules in step (a)/(aa) give rise to a replicable genetic package (RGP) displaying said (poly)peptides at its surface.

5. The method according to claim 4, wherein said recombinant vector molecule is a recombinant phage, phagemid or virus.

6. The method according to claim 5, wherein said phage is
(a) one of the class I phage fd, M13, If, Ike, ZJ/2, Ff;
(b) one of the class II phage Xf, Pf1, and Pf3;
(c) one of the lambdoid phages, lambda, 434, P1;
(d) one of the class of enveloped phages, PRD1; or
(e) one of the class paramyxo-viruses, orthomyxo-viruses, baculo-viruses, retro-viruses, reo-viruses and alpha-viruses.

7. The method according to any one of claims 4 to 6, wherein said selection step (e)/(b) is carried out by selecting polyphage comprising the interacting (poly)peptides.

8. The method according to any one of claims 4 to 7, wherein said screenable or selectable property is connected to the infectivity of said RGP.

9. The method according to claim 8, wherein said RGP is encoded by said recombinant vector used in step (a)/(aa) and rendered non-infective and infectivity of said RGP is restored by interaction of said (poly)peptide of step (a)/(aa) with the (poly)peptide of step (b)/(ab) and/or (c)/(ac), said (poly)peptide of step (b)/(ab) and/or (c)/(ac) being fused to a domain that confers infectivity to said RGP.

10. The method according to claim 9, wherein said RGP is rendered non-infective by modification of a genetic sequence which encodes a surface protein necessary for the RGP's binding to and infection of a host cell.

11. The method according to any one of claims 1 to 3, wherein said recombinant vector molecules in step (a)/(aa) give rise to a fusion protein which is expressed on the surface of a cell, preferably a bacterium.

12. The method according to claim 11, wherein said bacterium is Neisseria gonorrhoe or E. coli and said fusion protein consists of at least a part of a flagellum, lam B, peptidoglycan-associated lipoprotein or the Omp A protein and said (poly)peptide.

13. The method according to any one of claims 3 to 7, 11 or 12, wherein said (poly)peptides encoded by said recombinant vector molecules of step (b)/(ab) or (c)/(ac) are linked to at least one screenable or selectable tag.

14. The method according to claim 13, wherein said screenable or selectable tag is encoded by said recombinant vector of step (b)/(ab) or (c)/(ac).

15. The method according to claim 13 or 14, wherein said screenable or selectable tag is selected from the list His(n), myc, FLAG, malE, thioredoxin, GST, streptavidin, beta-galactosidase, alkaline phosphatase, T7 gene 10, Strep-tag and calmodulin.

16. The method according to claim 13, wherein said screenable or selectable tag is encoded by the genome of the host cell.

17. The method according to any one of claims 1 to 16, wherein said (poly)peptides encoded by the nucleic acid sequences of said additional libraries of step (c)/(ac) cause the interaction of said (poly)peptides of steps (a)/(aa) and (b)/(ab) via phosphorylation, glycosylation, methylation, lipidation or farnesylation of at least one of said (poly)peptides of steps (a)/(aa) and (b)/(ab).

18. The method according to any of claims 1 to 10 and 13 to 17, wherein said host cells in step (d)/(a) are spatially addressable, and the nucleic acid sequences mentioned in step (g)/(d) are retrieved from the corresponding spatially addressable host cell.

19. The method according to claim 1 or 2, wherein said screenable or selectable property is expressed intracellularly.

20. The method according to claim 19, wherein said screenable or selectable property is the transactivation of transcription of a reporter gene such as beta-galactosidase, alkaline phosphatase or nutritional markers such as his3 and leu, or resistance genes giving resistance to an antibiotic such as ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline or streptomycin.

21. The method according to any one of claims 1 to 20, wherein said recombinant vectors of step (a)/(aa), (b)/(ab) and (c)/(ac) comprise recombination promoting sites and in said step (e)/(b) recombination events are selected for, wherein said nucleic acid sequences encoding said (poly)peptides of step (a)/(aa), said nucleic acid sequences encoding said (poly)peptides of step (b)/(ab) and optionally said nucleic acid sequences encoding said (poly)peptides of step (c)/(ac) are contained in the same vector.

22. The method according to claim 21, wherein said recombination events are mediated by the site-specific recombination mechanisms Cre-lox, attP-attB, Mu gin or yeast flp.

23. The method according to claim 21 wherein said recombination promotion sites are restriction enzyme recognition sites and said recombination event is achieved by cutting the recombinant vector molecules mentioned in step (a)/(aa), (b)/(ab) and optionally (c)/(ac) with at least two different restriction enzymes and effecting recombination of the nucleic acid sequences contained in said vectors by ligation.

24. The method according to any one of claims 1 to 23 wherein said identification of said nucleic acid sequences is effected after the selection of step (e)/(b) via PCR and preferably sequencing of said nucleic acid sequences after said PCR.

25. The method according to any one of claims 1 to 24, wherein said recombinant vectors of step (a)/(aa), (b)/(ab) and/or (c)/(ac) comprise at least one gene encoding a selection marker.

26. The method according to claim 25, wherein said selection marker is a resistance to an antibiotic, preferably to ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline or streptomycin.

27. The method according to any one of claims 1 to 26, wherein said host cells are F' and preferably E.coli XL-1 Blue, K91 or its derivatives, TG1, XL1kan or TOP10F.

28. The method according to any one of claims 3 to 18 and 21 to 27, wherein said RGPs are produced with the use of helper phage taken from the list R408, M13k07 and VCSM13, M13de13, fCA55 and fKN16 or derivatives thereof.

29. The method according to any of claims 1 to 28, wherein at least one of said genetically diverse nucleic acid sequences encode members of the immunoglobulin superfamily.

30. The method according to claim 29, wherein said genetically diverse nucleic acid sequences encode a repertoire of immunoglobulin heavy or light chains.

31. The method according to any of claims 1 to 30, in which said genetically diverse nucleic acid sequences are generated by a mutagenesis method.

32. The method according to any of claims 1 to 31, in which said genetically diverse nucleic acid sequences are generated from a cDNA library.

33. The method according to any one of claims 1 to 32 wherein said nucleic acid sequences are genes or parts thereof.

34. Kit comprising at least

(a) a recombinant vector molecule as described in step (a)/(aa) or a corresponding vector molecule;

(b) a recombinant vector molecule as described in step (b)/(ab) or a corresponding vector molecule; and, optionally,

(c) at least one further recombinant vector molecule as described in step (c)/(ac) or a corresponding vector molecule.

Figure 1: General description of the polyphage principle

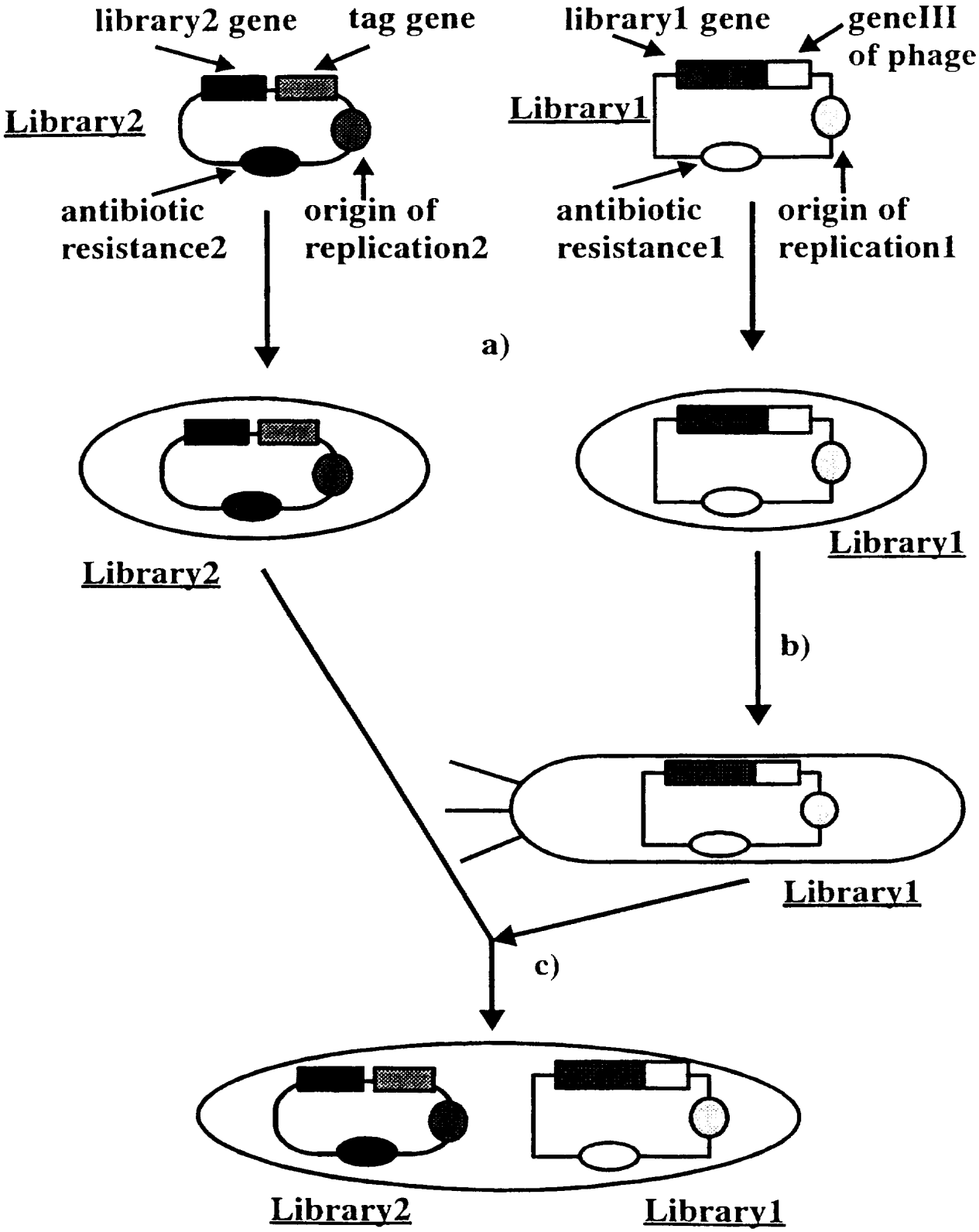


Figure 1: General description of the polyphage principle (cont.)

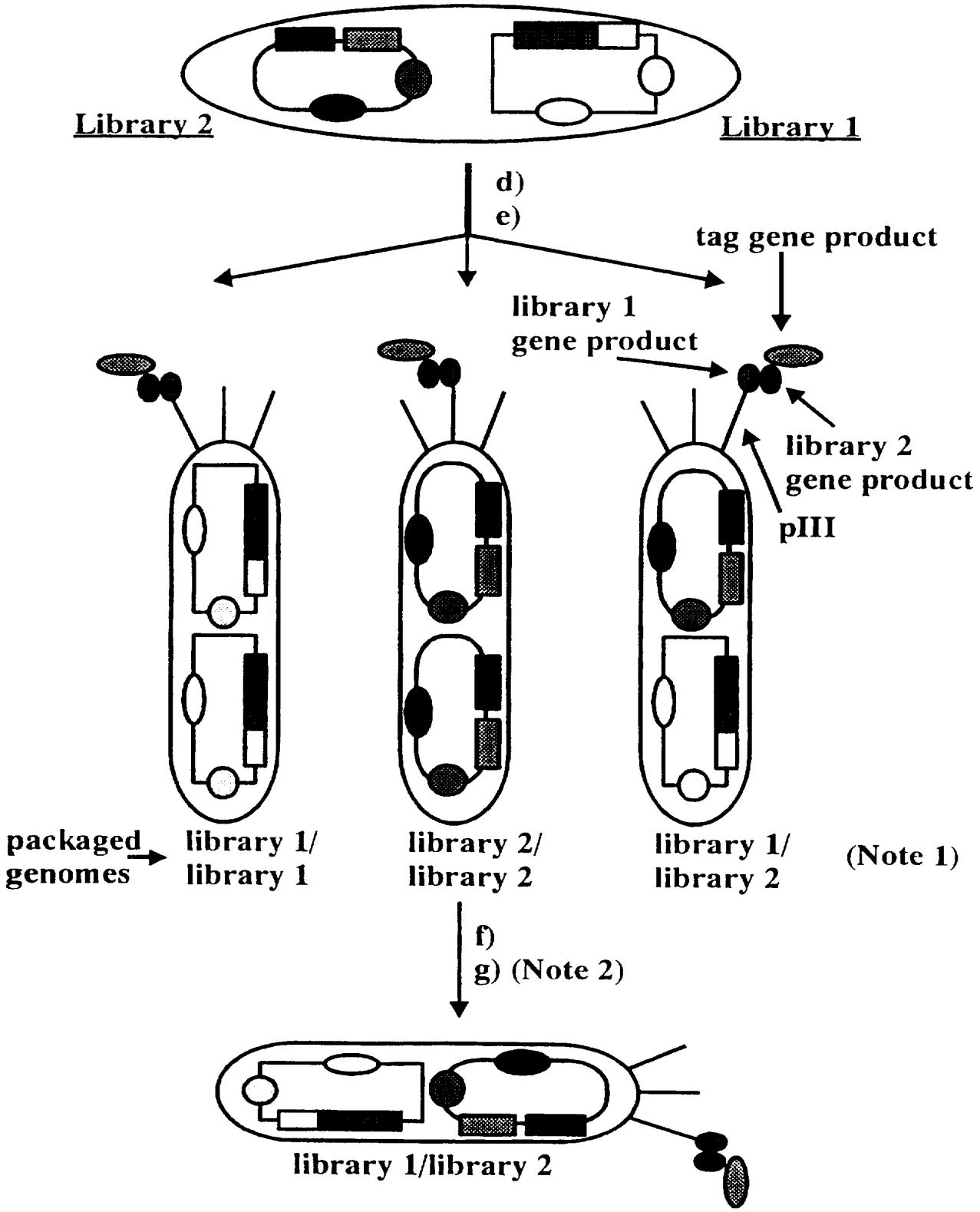


Figure 2: Co-transformation of two phagemids, polyphage formation and selection *via* His-tag: general description

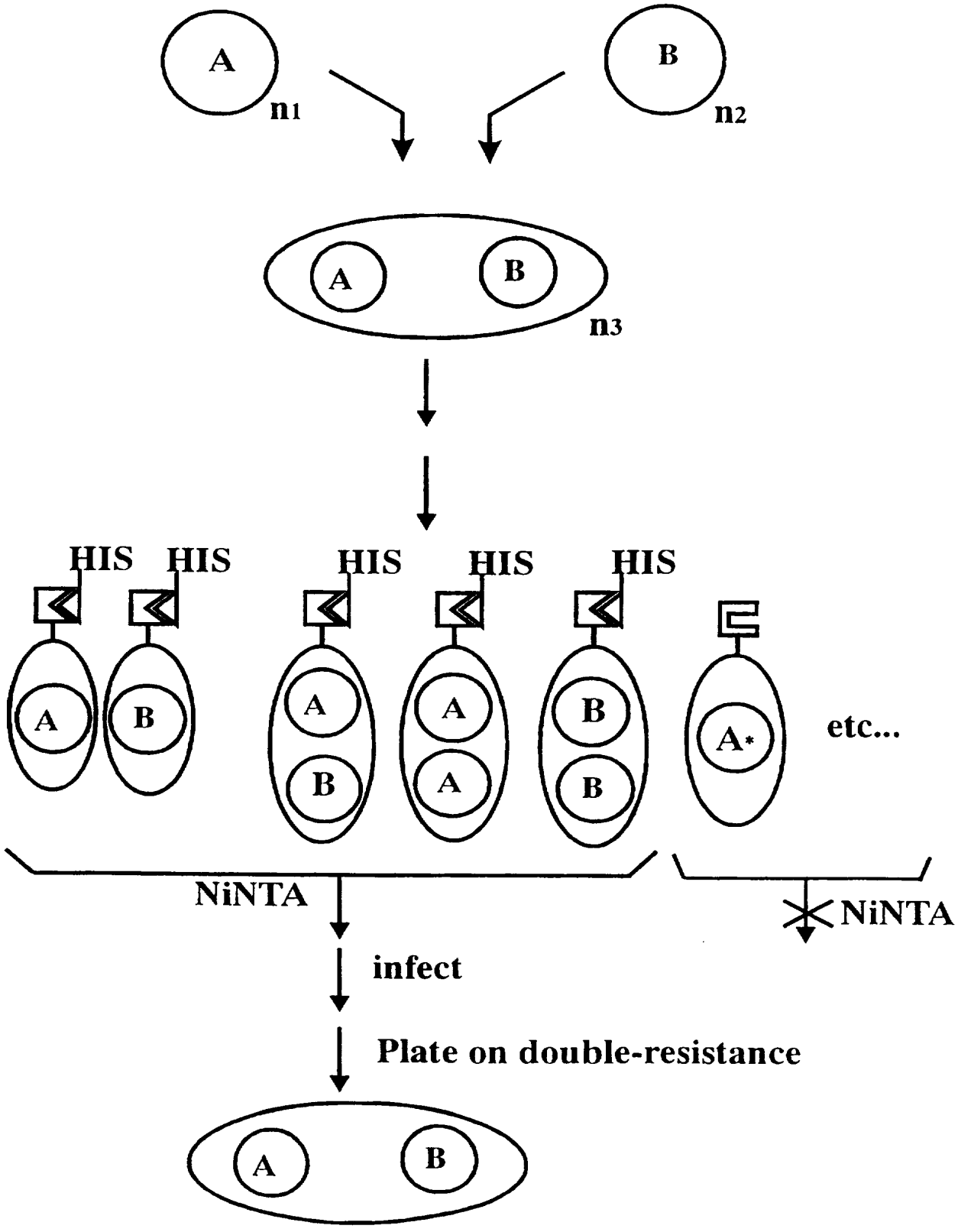


Figure 3: pBS vector series: functional map and sequence of pBS13

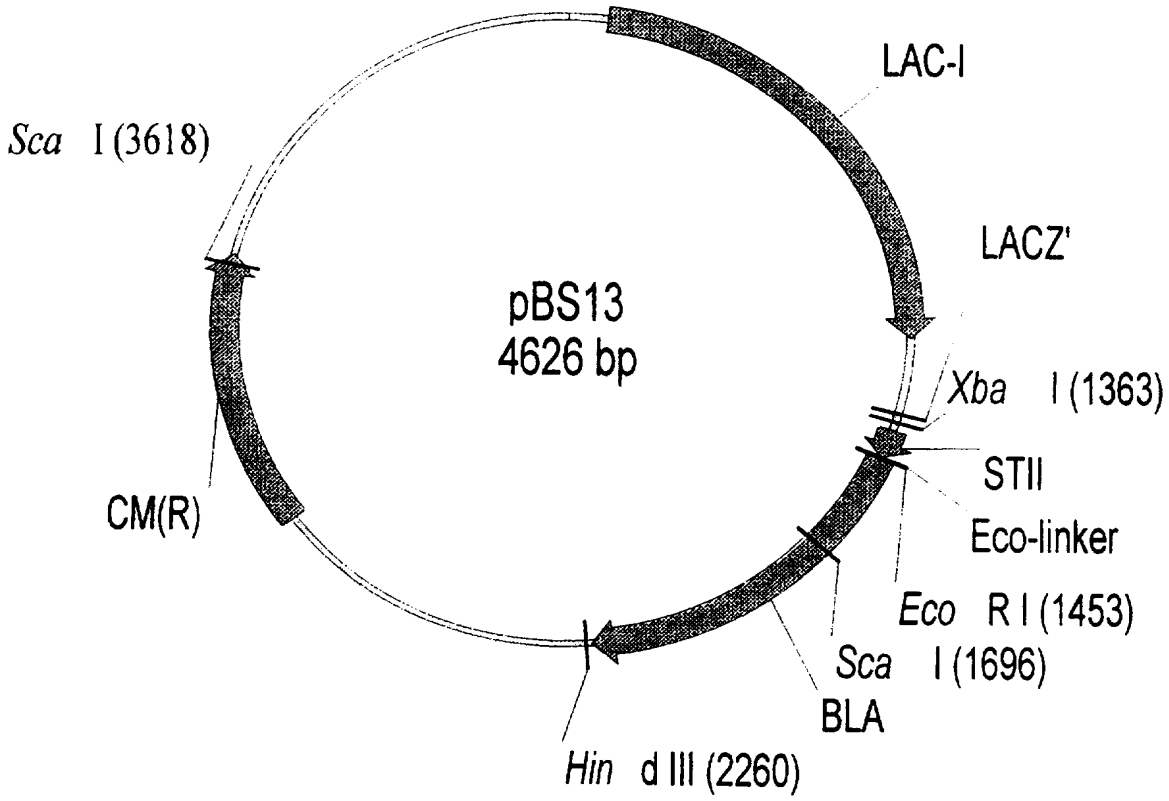


Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
   1  ACCCGACACC ATCGAATGGC GCAAAACCTT TCGCGGTATG GCATGATAGC
      TGGGCTGTGG TAGCTTACCG CGTTTTGGAA AGCGCCATAC CGTACTATCG

  51  GCCCGGAAGA GAGTCAATTC AGGGTGGTGA ATGTGAAACC AGTAACGTTA
      CGGGCCTTCT CTCAGTTAAG TCCCACCACT TACACTTTGG TCATTGCAAT

 101  TACGATGTCG CAGAGTATGC CGGTGTCTCT TATCAGACCG TTTCCCGCGT
      ATGCTACAGC GTCTCATACG GCCACAGAGA ATAGTCTGGC AAAGGGCGCA

 151  GGTGAACCAG GCCAGCCACG TTTCTGCGAA AACGCGGGAA AAAGTGGAAG
      CCACTTGGTC CGGTCGGTGC AAAGACGCTT TTGCGCCCTT TTTCACCTTC

 201  CGGCGATGGC GGAGCTGAAT TACATTCCCA ACCGCGTGGC ACAACAACTG
      GCCGCTACCG CCTCGACTTA ATGTAAGGGT TGGCGCACCG TGTTGTTGAC

 251  GCGGGCAAAC AGTCGTTGCT GATTGGCGTT GCCACCTCCA GTCTGGCCCT
      CGCCCGTTTG TCAGCAACGA CTAACCGCAA CGGTGGAGGT CAGACCGGGA

 301  GCACGCGCCG TCGCAAATTG TCGCGGCGAT TAAATCTCGC GCCGATCAAC
      CGTGCGCGGC AGCGTTTAAC AGCGCCGCTA ATTTAGAGCG CGGCTAGTTG

 351  TGGGTGCCAG CGTGGTGGTG TCGATGGTAG AACGAAGCGG CGTCGAAGCC
      ACCCACGGTC GCACCACCAC AGCTACCATC TTGCTTCGCC GCAGCTTCGG

 401  TGTAAAGCGG CGGTGCACAA TCTTCTCGCG CAACGCGTCA GTGGGCTGAT
      ACATTTCGCC GCCACGTGTT AGAAGAGCGC GTTGCGCAGT CACCCGACTA

 451  CATTAACTAT CCGCTGGATG ACCAGGATGC CATTGCTGTG GAAGCTGCCT
      GTAATTGATA GGCGACCTAC TGGTCCTACG GTAACGACAC CTTCGACGGA

 501  GCACTAATGT TCCGGCGTTA TTTCTTGATG TCTCTGACCA GACACCCATC
      CGTGATTACA AGGCCGCAAT AAAGAACTAC AGAGACTGGT CTGTGGGTAG

 551  AACAGTATTA TTTTCTCCCA TGAAGACGGT ACGCGACTGG GCGTGGAGCA
      TTGTCATAAT AAAAGAGGGT ACTTCTGCCA TGCGCTGACC CGCACCTCGT

 601  TCTGGTCGCA TTGGGTCACC AGCAAATCGC GCTGTTAGCG GGCCCATTAA
      AGACCAGCGT AACCCAGTGG TCGTTTAGCG CGACAATCGC CCGGGTAATT

 651  GTTCTGTCTC GGCGCGTCTG CGTCTGGCTG GCTGGCATAA ATATCTCACT
      CAAGACAGAG CCGCGCAGAC GCAGACCGAC CGACCGTATT TATAGAGTGA

 701  CGCAATCAAA TTCAGCCGAT AGCGGAACGG GAAGGCGACT GGAGTGCCAT
      GCGTTAGTTT AAGTCGGCTA TCGCCTTGCC CTTCCGCTGA CCTCACGGTA

 751  GTCCGGTTTT CAACAAACCA TGCAAATGCT GAATGAGGGC ATCGTTCCCA
      CAGGCCAAAA GTTGTTTGGT ACGTTTACGA CTTACTCCCG TAGCAAGGGT

 801  CTGCGATGCT GGTTGCCAAC GATCAGATGG CGCTGGGCGC AATGCGCGCC
      GACGCTACGA CCAACGGTTG CTAGTCTACC GCGACCCGCG TTACGCGCGG
```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
 851  ATTACCGAGT CCGGGCTGCG CGTTGGTGCG GACATCTCGG TAGTGGGATA
      TAATGGCTCA GGCCCGACGC GCAACCACGC CTGTAGAGCC ATCACCCTAT

 901  CGACGATACC GAAGACAGCT CATGTTATAT CCCGCCGTTA ACCACCATCA
      GCTGCTATGG CTTCTGTCGA GTACAATATA GGGCGGCAAT TGGTGGTAGT

 951  AACAGGATTT TCGCCTGCTG GGGCAAACCA GCGTGGACCG CTTGCTGCAA
      TTGTCCTAAA AGCGGACGAC CCCGTTTGGT CGCACCTGGC GAACGACGTT

1001  CTCTCTCAGG GCCAGGCGGT GAAGGGCAAT CAGCTGTTGC CCGTCTCACT
      GAGAGAGTCC CGGTCCGCCA CTTCCCGTTA GTCGACAACG GGCAGAGTGA

1051  GGTGAAAAGA AAAACCACCC TGGCGCCCAA TACGCAAACC GCCTCTCCCC
      CCACTTTTCT TTTTGGTGGG ACCGCGGGTT ATGCGTTTGG CGGAGAGGGG

1101  GCGCGTTGGC CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG
      CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGGCTGAC

1151  GAAAGCGGGC AGTGAGCGGT ACCCGATAAA AGCGGCTTCC TGACAGGAGG
      CTTTCGCCCG TCACTCGCCA TGGGCTATTT TCGCCGAAGG ACTGTCCTCC

1201  CCGTTTTGTT TTGCAGCCCA CCTCAACGCA ATTAATGTGA GTTAGCTCAC
      GGCAAAACAA AACGTCGGGT GGAGTTGCGT TAATTACACT CAATCGAGTG

1251  TCATTAGGCA CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT
      AGTAATCCGT GGGGTCCGAA ATGTGAAATA CGAAGGCCGA GCATACAACA

1301  GTGGAATTGT GAGCGGATAA CAATTTCACA CAGGAAACAG CTATGACCAT
      CACCTTAACA CTCGCCTATT GTTAAAGTGT GTCCTTTGTC GATACTGGTA

                  XbaI
                  ~~~~~~
1351  GATTACGAAT TTCTAGAGGT TGAGGTGATT TTATGAAAAA GAATATCGCA
      CTAATGCTTA AAGATCTCCA ACTCCACTAA AATACTTTTT CTTATAGCGT

1401  TTTCTTCTTG CATCTATGTT CGTTTTTTCT ATTGCTACAA ATGCATACGC
      AAAGAAGAAC GTAGATACAA GCAAAAAAGA TAACGATGTT TACGTATGCG

        EcoRI
        ~~~~~~
1451  TGAATTCCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT
      ACTTAAGGTG GGTCTTTGCG ACCACTTTCA TTTTCTACGA CTTCTAGTCA

1501  TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC
      ACCCACGTGC TCACCCAATG TAGCTTGACC TAGAGTTGTC GCCATTCTAG

1551  CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA
      GAACTCTCAA AAGCGGGGCT TCTTGCAAAA GGTTACTACT CGTGAAAATT

1601  AGTTCTGCTA TGTGGCGCGG TATTATCCCG TATTGACGCC GGGCAAGAGC
      TCAAGACGAT ACACCGCGCC ATAATAGGGC ATAACTGCGG CCCGTTCTCG
```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
                                                                  ScaI
                                                                  ~~~~~~
1651  AACTCGGTCG CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA
      TTGAGCCAGC GGCGTATGTG ATAAGAGTCT TACTGAACCA ACTCATGAGT

1701  CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA GAGAATTATG
      GGTCAGTGTC TTTTCGTAGA ATGCCTACCG TACTGTCATT CTCTTAATAC

1751  CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA
      GTCACGACGG TATTGGTACT CACTATTGTG ACGCCGGTTG AATGAAGACT

1801  CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG
      GTTGCTAGCC TCCTGGCTTC CTCGATTGGC GAAAAAACGT GTTGTACCCC

1851  GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT
      CTAGTACATT GAGCGGAACT AGCAACCCTT GGCCTCGACT TACTTCGGTA

1901  ACCAAACGAC GAGCGTGACA CCACGATGCC TGTAGCAATG GCAACAACGT
      TGGTTTGCTG CTCGCACTGT GGTGCTACGG ACATCGTTAC CGTTGTTGCA

1951  TGCGCAAACT ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA
      ACGCGTTTGA TAATTGACCG CTTGATGAAT GAGATCGAAG GGCCGTTGTT

2001  TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC
      AATTATCTGA CCTACCTCCG CCTATTTCAA CGTCCTGGTG AAGACGCGAG

2051  GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC
      CCGGGAAGGC CGACCGACCA AATAACGACT ATTTAGACCT CGGCCACTCG

2101  GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC
      CACCCAGAGC GCCATAGTAA CGTCGTGACC CCGGTCTACC ATTCGGGAGG

2151  CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG
      GCATAGCATC AATAGATGTG CTGCCCCTCA GTCCGTTGAT ACCTACTTGC

2201  AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAT
      TTTATCTGTC TAGCGACTCT ATCCACGGAG TGACTAATTC GTAACCATTA

            HindIII
            ~~~~~~~
2251  GAGCATGCAA GCTTGACCTG TGAAGTGAAA AATGGCGCAC ATTGTGCGAC
      CTCGTACGTT CGAACTGGAC ACTTCACTTT TTACCGCGTG TAACACGCTG

2301  ATTTTTTTTG TCTGCCGTTT ACCGCTACTG CGTCACGGAT CCCCACGCGC
      TAAAAAAAAC AGACGGCAAA TGGCGATGAC GCAGTGCCTA GGGGTGCGCG

2351  CCTGTAGCGG CGCATTAAGC GCGGCGGGTG TGGTGGTTAC GCGCAGCGTG
      GGACATCGCC GCGTAATTCG CGCCGCCCAC ACCACCAATG CGCGTCGCAC

2401  ACCGCTACAC TTGCCAGCGC CCTAGCGCCC GCTCCTTTCG CTTTCTTCCC
      TGGCGATGTG AACGGTCGCG GGATCGCGGG CGAGGAAAGC GAAAGAAGGG
```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
2451  TTCCTTTCTC GCCACGTTCG CCGGCTTTCC CCGTCAAGCT CTAAATCGGG
      AAGGAAAGAG CGGTGCAAGC GGCCGAAAGG GGCAGTTCGA GATTTAGCCC

2501  GCATCCCTTT AGGGTTCCGA TTTAGTGCTT TACGGCACCT CGACCCCAAA
      CGTAGGGAAA TCCCAAGGCT AAATCACGAA ATGCCGTGGA GCTGGGGTTT

2551  AAACTTGATT AGGGTGATGG TTCACGTAGT GGGCCATCGC CCTGATAGAC
      TTTGAACTAA TCCCACTACC AAGTGCATCA CCCGGTAGCG GGACTATCTG

2601  GGTTTTTCGC CCTTTGACGT TGGAGTCCAC GTTCTTTAAT AGTGGACTCT
      CCAAAAAGCG GGAAACTGCA ACCTCAGGTG CAAGAAATTA TCACCTGAGA

2651  TGTTCCAAAC TGGAACAACA CTCAACCCTA TCTCGGTCTA TTCTTTTGAT
      ACAAGGTTTG ACCTTGTTGT GAGTTGGGAT AGAGCCAGAT AAGAAAACTA

2701  TTATAAGGGA TTTTGCCGAT TTCGGCCTAT TGGTTAAAAA ATGAGCTGAT
      AATATTCCCT AAAACGGCTA AAGCCGGATA ACCAATTTTT TACTCGACTA

2751  TTAACAAAAA TTTAACGCGA ATTTTAACAA AATATTAACG TTTACAATTT
      AATTGTTTTT AAATTGCGCT TAAAATTGTT TTATAATTGC AAATGTTAAA

2801  CAGGTGGCAC TTTTCGGGGA AATGTGCGCG GAACCCCTAT TTGTTTATTT
      GTCCACCGTG AAAAGCCCCT TTACACGCGC CTTGGGGATA AACAAATAAA

2851  TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGTCGAGAC GTTGGGTGAG
      AAGATTTATG TAAGTTTATA CATAGGCGAG TACAGCTCTG CAACCCACTC

2901  GTTCCAACTT TCACCATAAT GAAATAAGAT CACTACCGGG CGTATTTTTT
      CAAGGTTGAA AGTGGTATTA CTTTATTCTA GTGATGGCCC GCATAAAAAA

2951  GAGTTATCGA GATTTTCAGG AGCTAAGGAA GCTAAAATGG AGAAAAAAAT
      CTCAATAGCT CTAAAAGTCC TCGATTCCTT CGATTTTACC TCTTTTTTTA

3001  CACTGGATAT ACCACCGTTG ATATATCCCA ATGGCATCGT AAAGAACATT
      GTGACCTATA TGGTGGCAAC TATATAGGGT TACCGTAGCA TTTCTTGTAA

3051  TTGAGGCATT TCAGTCAGTT GCTCAATGTA CCTATAACCA GACCGTTCAG
      AACTCCGTAA AGTCAGTCAA CGAGTTACAT GGATATTGGT CTGGCAAGTC

3101  CTGGATATTA CGGCCTTTTT AAAGACCGTA AAGAAAAATA AGCACAAGTT
      GACCTATAAT GCCGGAAAAA TTTCTGGCAT TTCTTTTTAT TCGTGTTCAA

3151  TTATCCGGCC TTTATTCACA TTCTTGCCCG CCTGATGAAT GCTCATCCGG
      AATAGGCCGG AAATAAGTGT AAGAACGGGC GGACTACTTA CGAGTAGGCC

3201  AGTTCCGTAT GGCAATGAAA GACGGTGAGC TGGTGATATG GGATAGTGTT
      TCAAGGCATA CCGTTACTTT CTGCCACTCG ACCACTATAC CCTATCACAA

3251  CACCCTTGTT ACACCGTTTT CCATGAGCAA ACTGAAACGT TTTCATCGCT
      GTGGGAACAA TGTGGCAAAA GGTACTCGTT TGACTTTGCA AAAGTAGCGA
```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
3301  CTGGAGTGAA TACCACGACG ATTTCCGGCA GTTTCTACAC ATATATTCGC
      GACCTCACTT ATGGTGCTGC TAAAGGCCGT CAAAGATGTG TATATAAGCG

3351  AAGATGTGGC GTGTTACGGT GAAAACCTGG CCTATTTCCC TAAAGGGTTT
      TTCTACACCG CACAATGCCA CTTTTGGACC GGATAAAGGG ATTTCCCAAA

3401  ATTGAGAATA TGTTTTTCGT CTCAGCCAAT CCCTGGGTGA GTTTCACCAG
      TAACTCTTAT ACAAAAAGCA GAGTCGGTTA GGGACCCACT CAAAGTGGTC

3451  TTTTGATTTA AACGTGGCCA ATATGGACAA CTTCTTCGCC CCCGTTTTCA
      AAAACTAAAT TTGCACCGGT TATACCTGTT GAAGAAGCGG GGGCAAAAGT

3501  CCATGGGCAA ATATTATACG CAAGGCGACA AGGTGCTGAT GCCGCTGGCG
      GGTACCCGTT TATAATATGC GTTCCGCTGT TCCACGACTA CGGCGACCGC

3551  ATTCAGGTTC ATCATGCCGT CTGTGATGGC TTCCATGTCG GCAGAATGCT
      TAAGTCCAAG TAGTACGGCA GACACTACCG AAGGTACAGC CGTCTTACGA

                      ScaI
                      ~~~~~~
3601  TAATGAATTA CAACAGTACT GCGATGAGTG GCAGGGCGGG GCGTAATTTT
      ATTACTTAAT GTTGTCATGA CGCTACTCAC CGTCCCGCCC CGCATTAAAA

3651  TTTAAGGCAG TTATTGGTGC CCTTAAACGC CTGGTGCTAC GCCTGAATAA
      AAATTCCGTC AATAACCACG GGAATTTGCG GACCACGATG CGGACTTATT

3701  GTGATAATAA GCGGATGAAT GGCAGAAATT CGAAAGCAAA TTCGACCCGG
      CACTATTATT CGCCTACTTA CCGTCTTTAA GCTTTCGTTT AAGCTGGGCC

3751  TCGTCGGTTC AGGGCAGGGT CGTTAAATAG CCGCTTATGT CTATTGCTGG
      AGCAGCCAAG TCCCGTCCCA GCAATTTATC GGCGAATACA GATAACGACC

3801  TTTACCGGTT TATTGACTAC CGGAAGCAGT GTGACCGTGT GCTTCTCAAA
      AAATGGCCAA ATAACTGATG GCCTTCGTCA CACTGGCACA CGAAGAGTTT

3851  TGCCTGAGGC CAGTTTGCTC AGGCTCTCCC CGTGGAGGTA ATAATTGCTC
      ACGGACTCCG GTCAAACGAG TCCGAGAGGG GCACCTCCAT TATTAACGAG

3901  GACATGACCA AAATCCCTTA ACGTGAGTTT TCGTTCCACT GAGCGTCAGA
      CTGTACTGGT TTTAGGGAAT TGCACTCAAA AGCAAGGTGA CTCGCAGTCT

3951  CCCCGTAGAA AAGATCAAAG GATCTTCTTG AGATCCTTTT TTTCTGCGCG
      GGGGCATCTT TTCTAGTTTC CTAGAAGAAC TCTAGGAAAA AAAGACGCGC

4001  TAATCTGCTG CTTGCAAACA AAAAAACCAC CGCTACCAGC GGTGGTTTGT
      ATTAGACGAC GAACGTTTGT TTTTTTGGTG GCGATGGTCG CCACCAAACA

4051  TTGCCGGATC AAGAGCTACC AACTCTTTTT CCGAAGGTAA CTGGCTTCAG
      AACGGCCTAG TTCTCGATGG TTGAGAAAAA GGCTTCCATT GACCGAAGTC
```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
4101  CAGAGCGCAG ATACCAAATA CTGTCCTTCT AGTGTAGCCG TAGTTAGGCC
      GTCTCGCGTC TATGGTTTAT GACAGGAAGA TCACATCGGC ATCAATCCGG

4151  ACCACTTCAA GAACTCTGTA GCACCGCCTA CATACCTCGC TCTGCTAATC
      TGGTGAAGTT CTTGAGACAT CGTGGCGGAT GTATGGAGCG AGACGATTAG

4201  CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT
      GACAATGGTC ACCGACGACG GTCACCGCTA TTCAGCACAG AATGGCCCAA

4251  GGACTCAAGA CGATAGTTAC CGGATAAGGC GCAGCGGTCG GGCTGAACGG
      CCTGAGTTCT GCTATCAATG GCCTATTCCG CGTCGCCAGC CCGACTTGCC

4301  GGGGTTCGTG CACACAGCCC AGCTTGGAGC GAACGACCTA CACCGAACTG
      CCCCAAGCAC GTGTGTCGGG TCGAACCTCG CTTGCTGGAT GTGGCTTGAC

4351  AGATACCTAC AGCGTGAGCT ATGAGAAAGC GCCACGCTTC CCGAAGGGAG
      TCTATGGATG TCGCACTCGA TACTCTTTCG CGGTGCGAAG GGCTTCCCTC

4401  AAAGGCGGAC AGGTATCCGG TAAGCGGCAG GGTCGGAACA GGAGAGCGCA
      TTTCCGCCTG TCCATAGGCC ATTCGCCGTC CCAGCCTTGT CCTCTCGCGT

4451  CGAGGGAGCT TCCAGGGGGA AACGCCTGGT ATCTTTATAG TCCTGTCGGG
      GCTCCCTCGA AGGTCCCCCT TTGCGGACCA TAGAAATATC AGGACAGCCC

4501  TTTCGCCACC TCTGACTTGA GCGTCGATTT TTGTGATGCT CGTCAGGGGG
      AAAGCGGTGG AGACTGAACT CGCAGCTAAA AACACTACGA GCAGTCCCCC

4551  GCGGAGCCTA TGGAAAAACG CCAGCAACGC GGCCTTTTTA CGGTTCCTGG
      CGCCTCGGAT ACCTTTTTGC GGTCGTTGCG CCGGAAAAAT GCCAAGGACC

4601  CCTTTTGCTG GCCTTTTGCT CACATG
      GGAAAACGAC CGGAAAACGA GTGTAC
```

Figure 4: Co-existence of phagemids: results of restriction digest

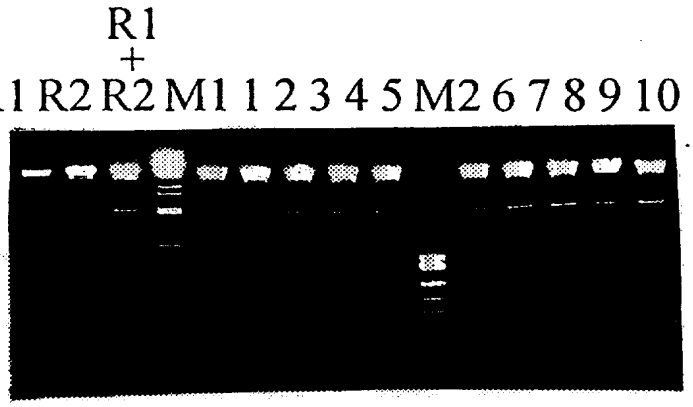


Figure 5: Phagemid vector pYING1-C1: functional map

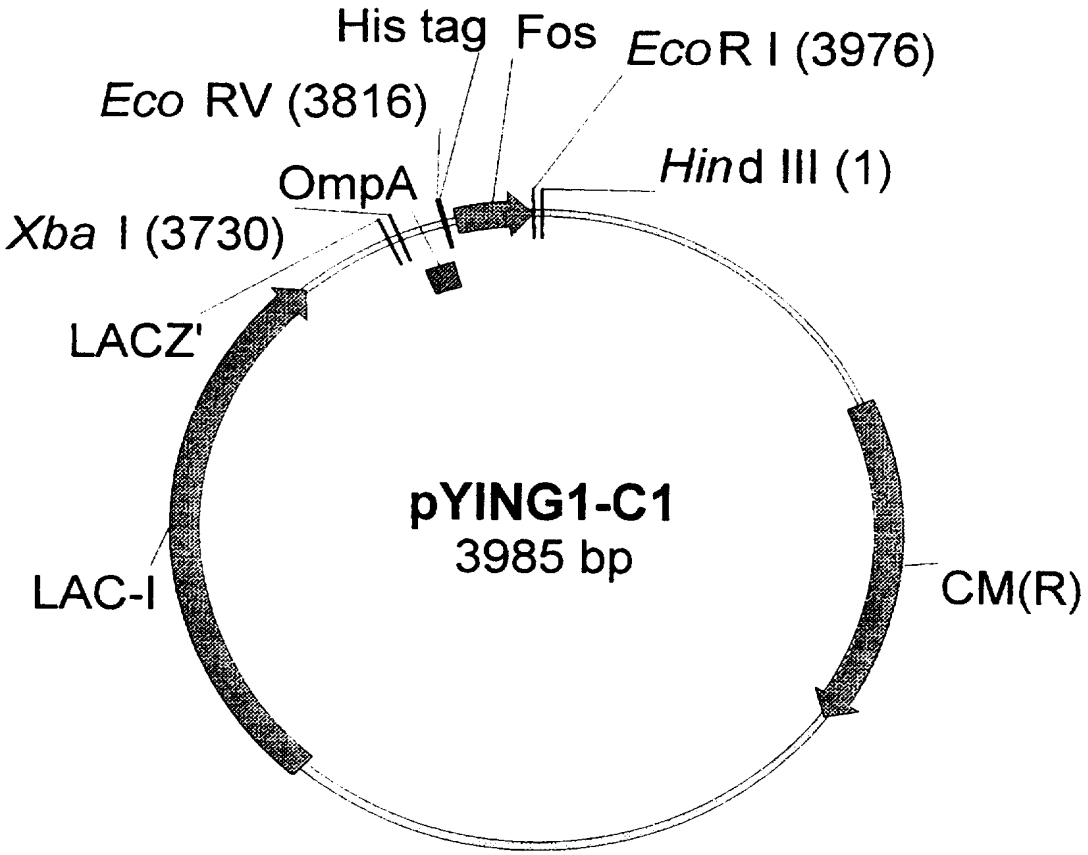


Figure 6: Phagemid vector pYANG3-A: functional map

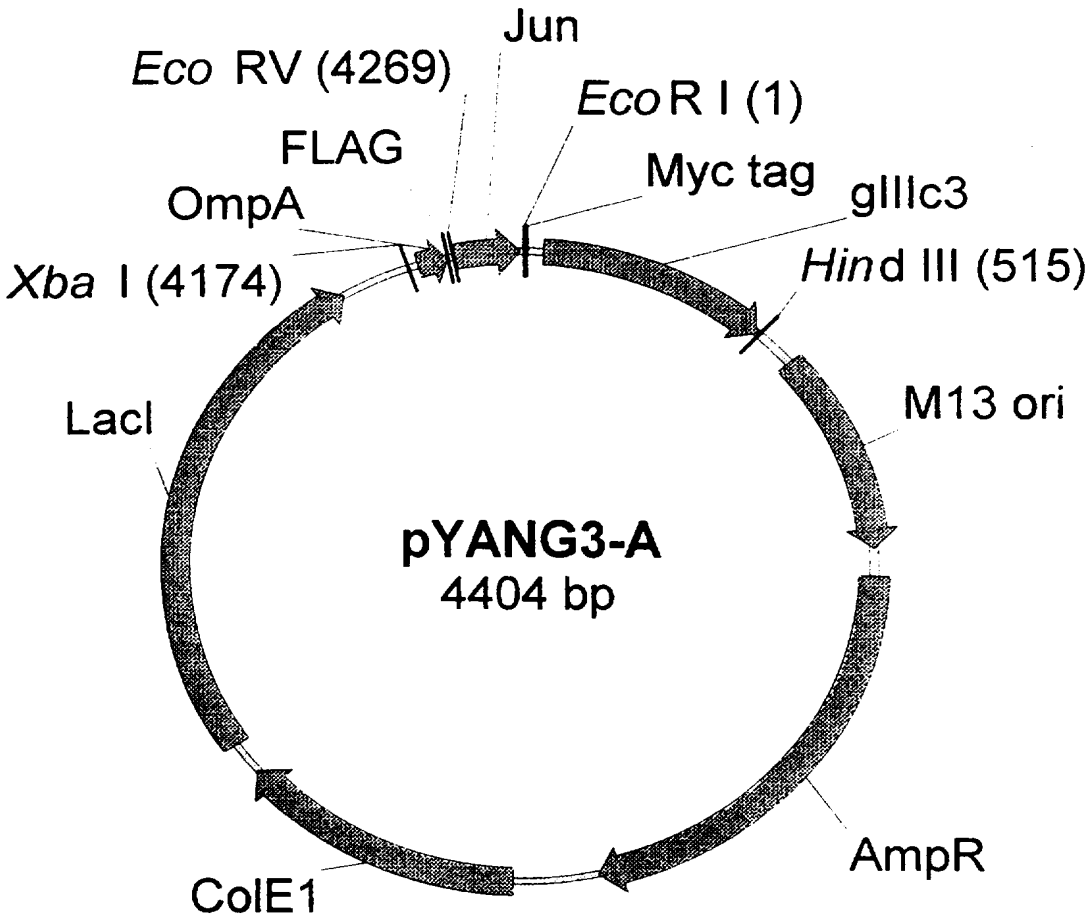
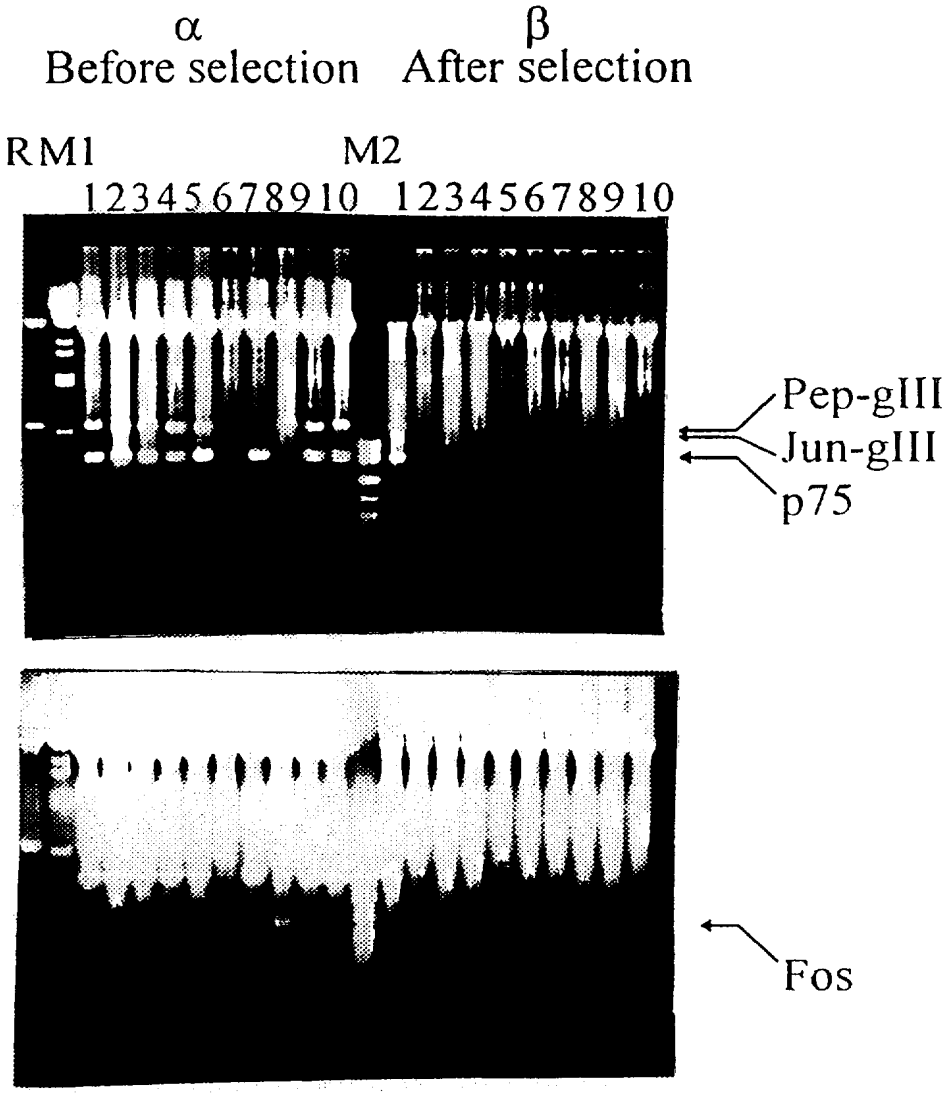


Figure 7: Analysis of selected clones (see Table 2)

7.a: Restriction digest of clones before and after selection



7.b: PCR of clones after selection with primers OPEP5L and OGIII3

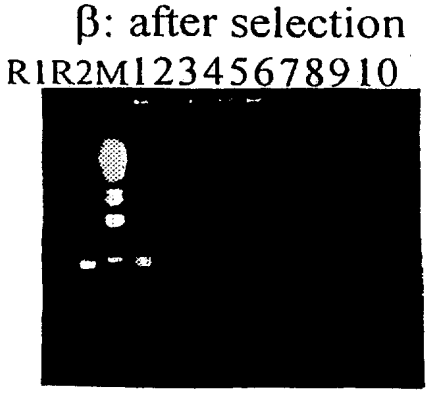


Figure 8: Phagemid vector pING1-C1: functional map

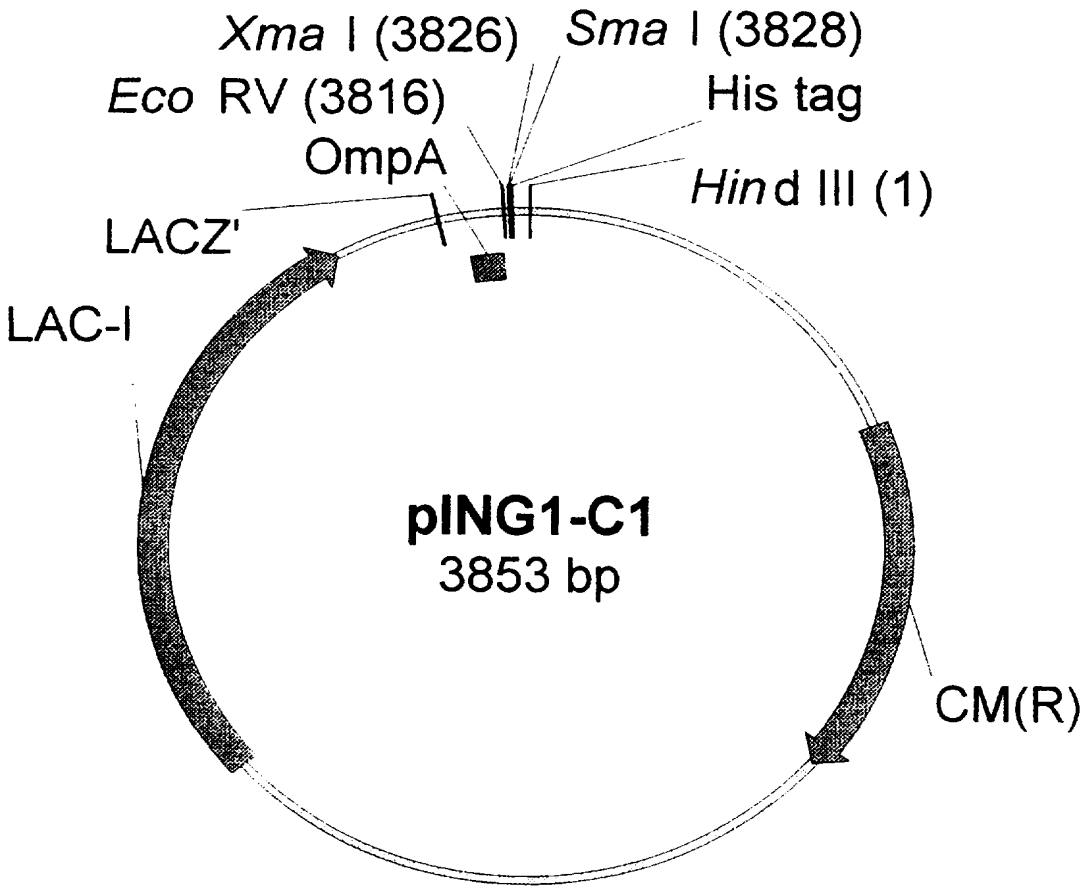


Figure 9: Phagemid vector pONG3-A: functional map

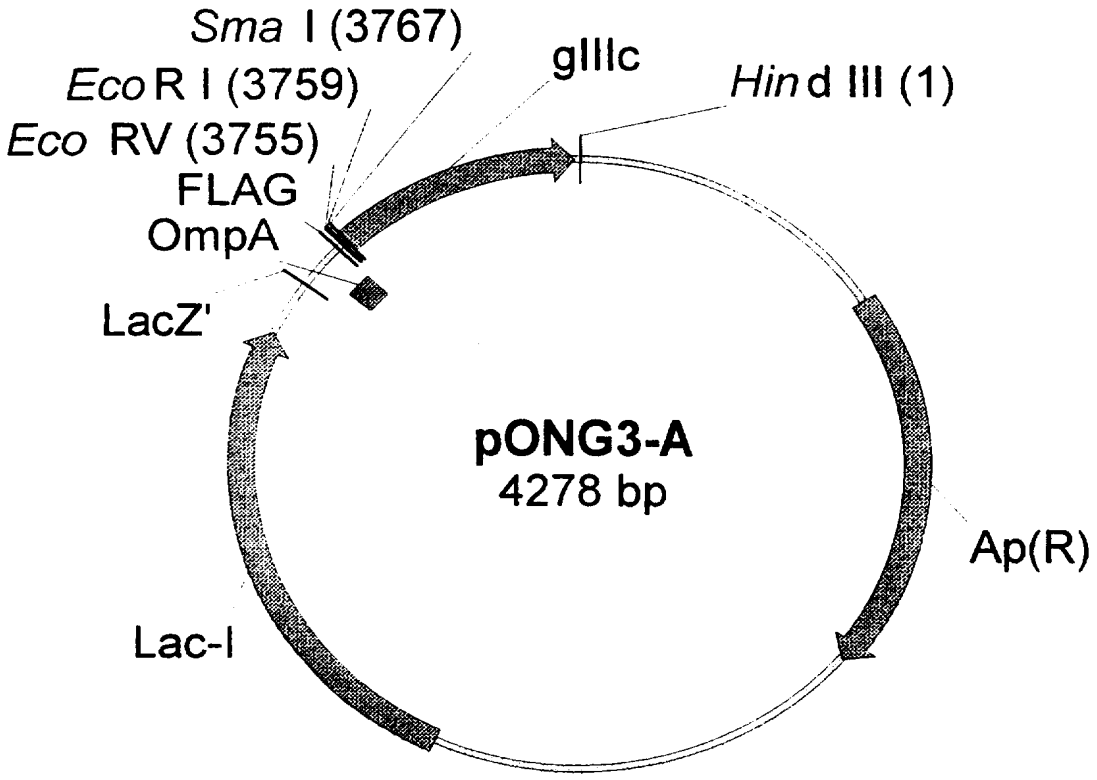


Figure 10: Co-transformation of phage and plasmid, polyphage formation and selection *via* SIP: general description

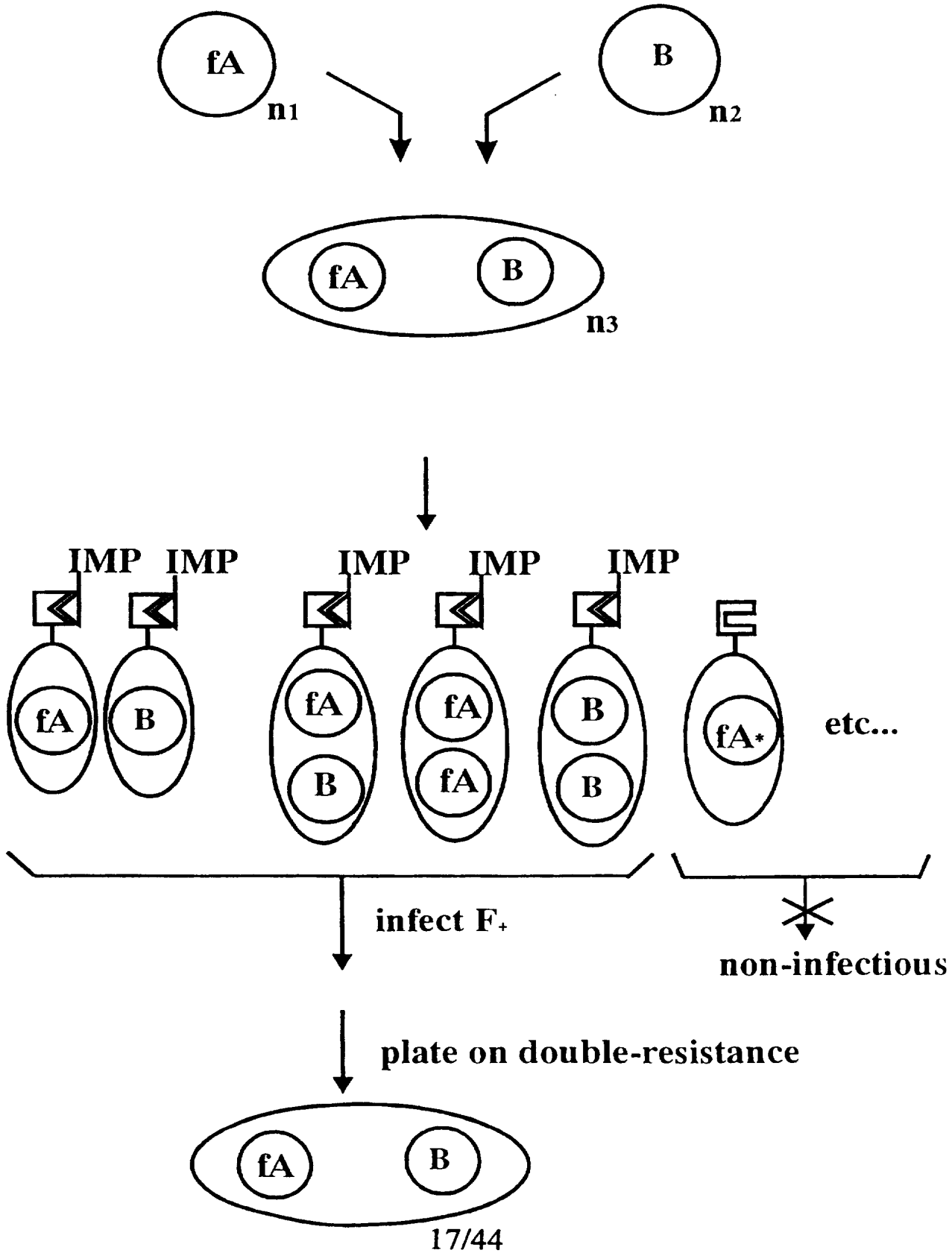


Figure 11: Phage vector fhag1A: functional map

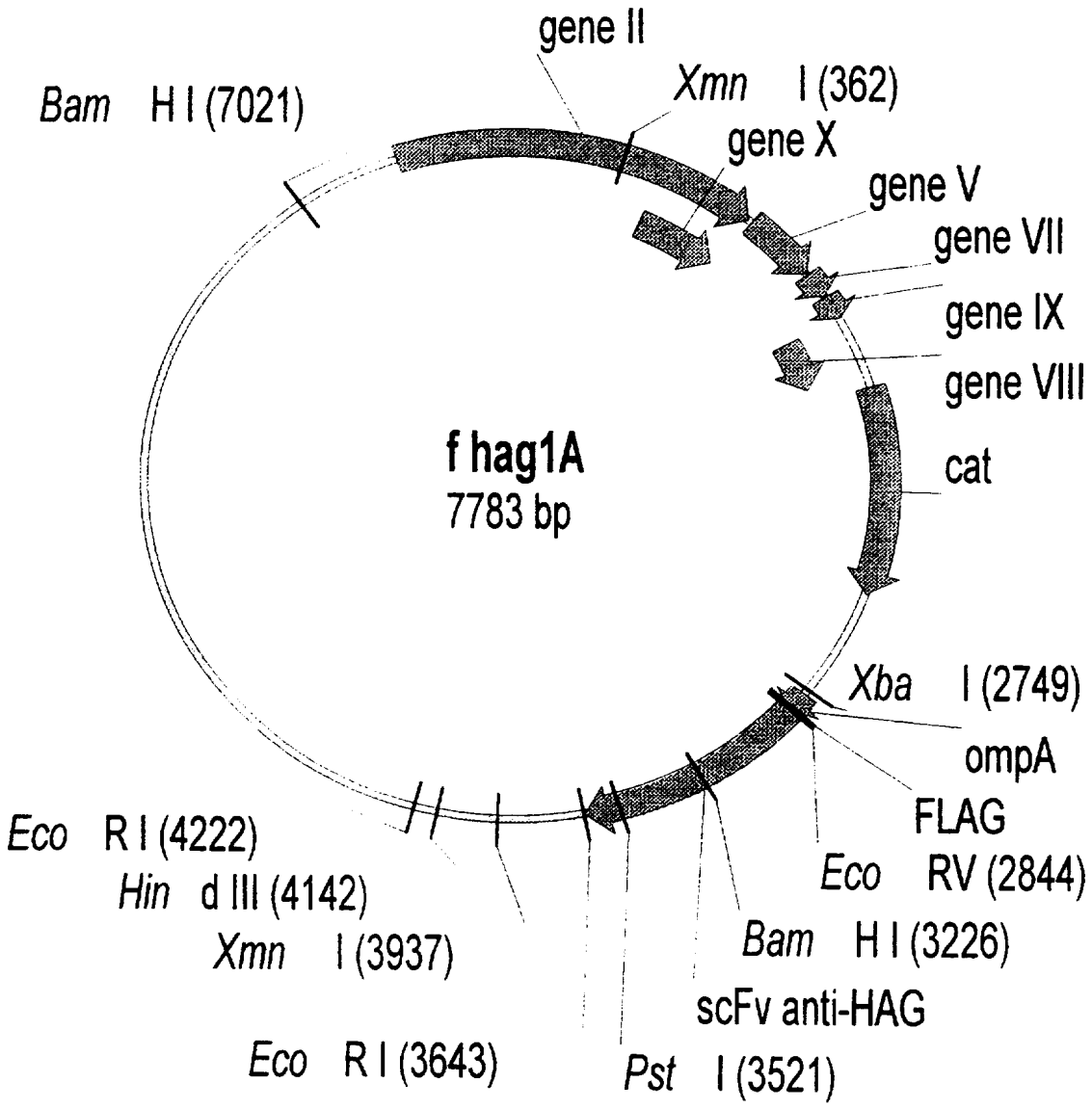


Figure 11a: CAT gene module: functional map and sequence

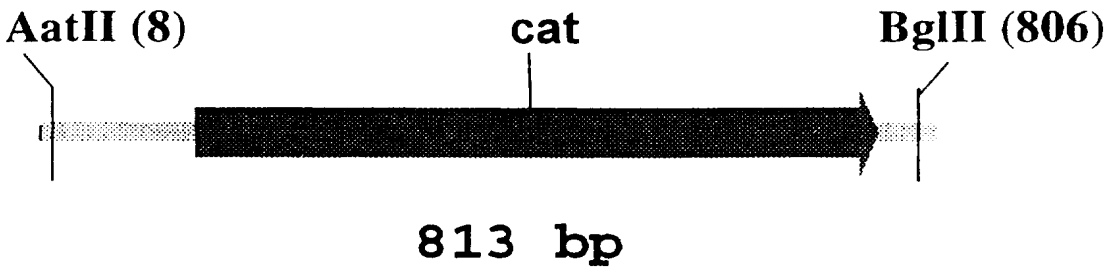


Figure 11a: CAT gene module: functional map and sequence (cont.)

```
          AatII
          ~~~~~~
    1  GGGACGTCGG GTGAGGTTCC AACTTTCACC ATAATGAAAT AAGATCACTA
       CCCTGCAGCC CACTCCAAGG TTGAAAGTGG TATTACTTTA TTCTAGTGAT

   51  CCGGGCGTAT TTTTTGAGTT ATCGAGATTT TCAGGAGCTA AGGAAGCTAA
       GGCCCGCATA AAAAACTCAA TAGCTCTAAA AGTCCTCGAT TCCTTCGATT

  101  AATGGAGAAA AAAATCACTG GATATACCAC CGTTGATATA TCCCAATGGC
       TTACCTCTTT TTTTAGTGAC CTATATGGTG GCAACTATAT AGGGTTACCG

  151  ATCGTAAAGA ACATTTTGAG GCATTTCAGT CAGTTGCTCA ATGTACCTAT
       TAGCATTTCT TGTAAAACTC CGTAAAGTCA GTCAACGAGT TACATGGATA

  201  AACCAGACCG TTCAGCTGGA TATTACGGCC TTTTTAAAGA CCGTAAAGAA
       TTGGTCTGGC AAGTCGACCT ATAATGCCGG AAAAATTTCT GGCATTTCTT

  251  AAATAAGCAC AAGTTTTATC CGGCCTTTAT TCACATTCTT GCCCGCCTGA
       TTTATTCGTG TTCAAAATAG GCCGGAAATA AGTGTAAGAA CGGGCGGACT

  301  TGAATGCTCA CCCGGAGTTC CGTATGGCAA TGAAAGACGG TGAGCTGGTG
       ACTTACGAGT GGGCCTCAAG GCATACCGTT ACTTTCTGCC ACTCGACCAC

  351  ATATGGGATA GTGTTCACCC TTGTTACACC GTTTTCCATG AGCAAACTGA
       TATACCCTAT CACAAGTGGG AACAATGTGG CAAAAGGTAC TCGTTTGACT

  401  AACGTTTTCA TCGCTCTGGA GTGAATACCA CGACGATTTC CGGCAGTTTC
       TTGCAAAAGT AGCGAGACCT CACTTATGGT GCTGCTAAAG GCCGTCAAAG

  451  TACACATATA TTCGCAAGAT GTGGCGTGTT ACGGTGAAAA CCTGGCCTAT
       ATGTGTATAT AAGCGTTCTA CACCGCACAA TGCCACTTTT GGACCGGATA

  501  TTCCCTAAAG GGTTTATTGA GAATATGTTT TTCGTCTCAG CCAATCCCTG
       AAGGGATTTC CCAAATAACT CTTATACAAA AAGCAGAGTC GGTTAGGGAC

  551  GGTGAGTTTC ACCAGTTTTG ATTTAAACGT AGCCAATATG GACAACTTCT
       CCACTCAAAG TGGTCAAAAC TAAATTTGCA TCGGTTATAC CTGTTGAAGA

  601  TCGCCCCCGT TTTCACTATG GGCAAATATT ATACGCAAGG CGACAAGGTG
       AGCGGGGGCA AAAGTGATAC CCGTTTATAA TATGCGTTCC GCTGTTCCAC

  651  CTGATGCCGC TGGCGATTCA GGTTCATCAT GCCGTTTGTG ATGGCTTCCA
       GACTACGGCG ACCGCTAAGT CCAAGTAGTA CGGCAAACAC TACCGAAGGT

  701  TGTCGGCAGA ATGCTTAATG AATTACAACA GTACTGCGAT GAGTGGCAGG
       ACAGCCGTCT TACGAATTAC TTAATGTTGT CATGACGCTA CTCACCGTCC

  751  GCGGGGCGTA ATTTTTTTAA GGCAGTTATT GGGTGCCCTT AAACGCCTGG
       CGCCCCGCAT TAAAAAAATT CCGTCAATAA CCCACGGGAA TTTGCGGACC

          BglII
          ~~~~~~
  801  TGCTAGATCT TCC
       ACGATCTAGA AGG
```

Figure 12: Phage vector fjun1A: functional map

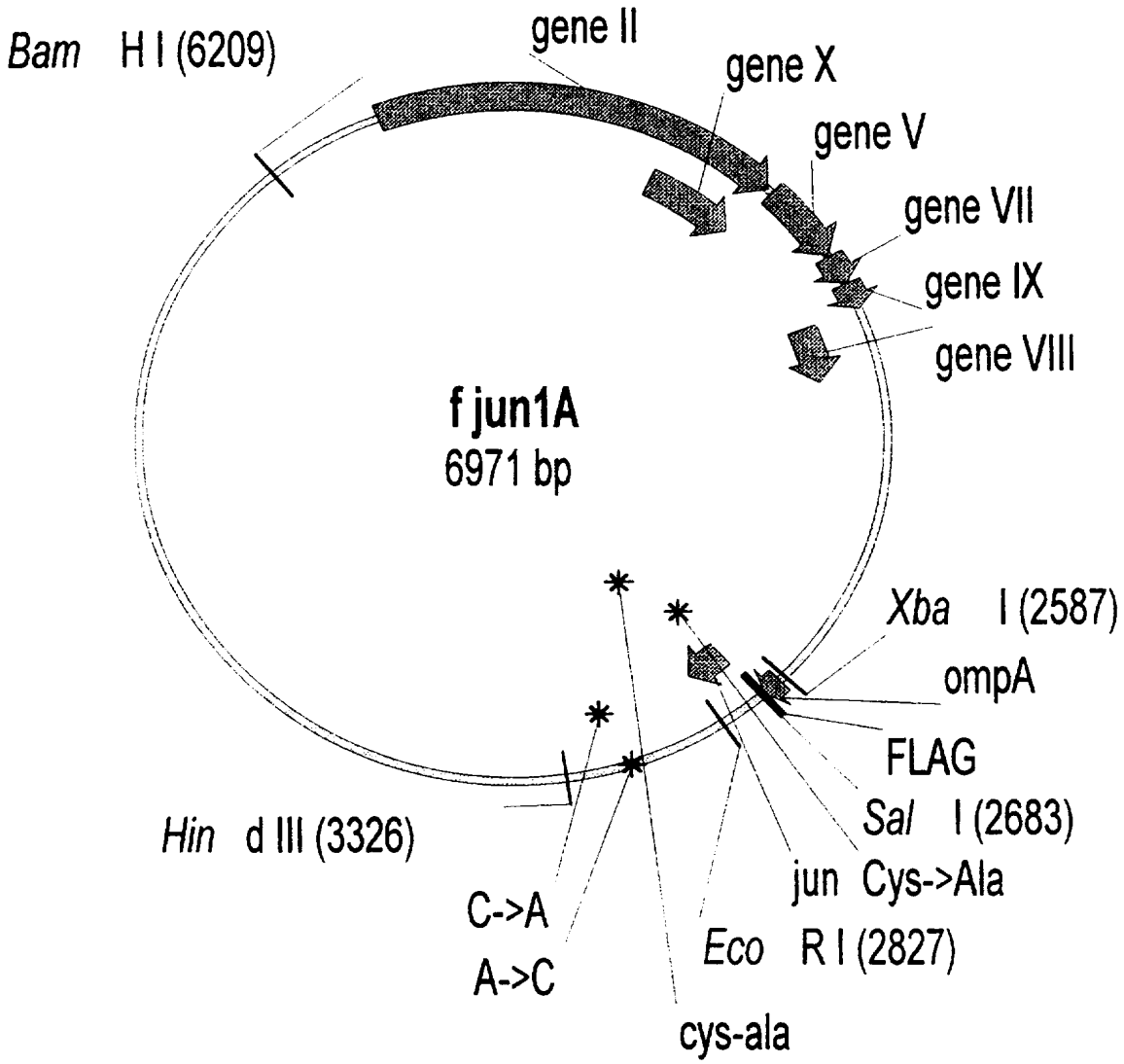


Figure 13: Phage vector fjun1B: functional map

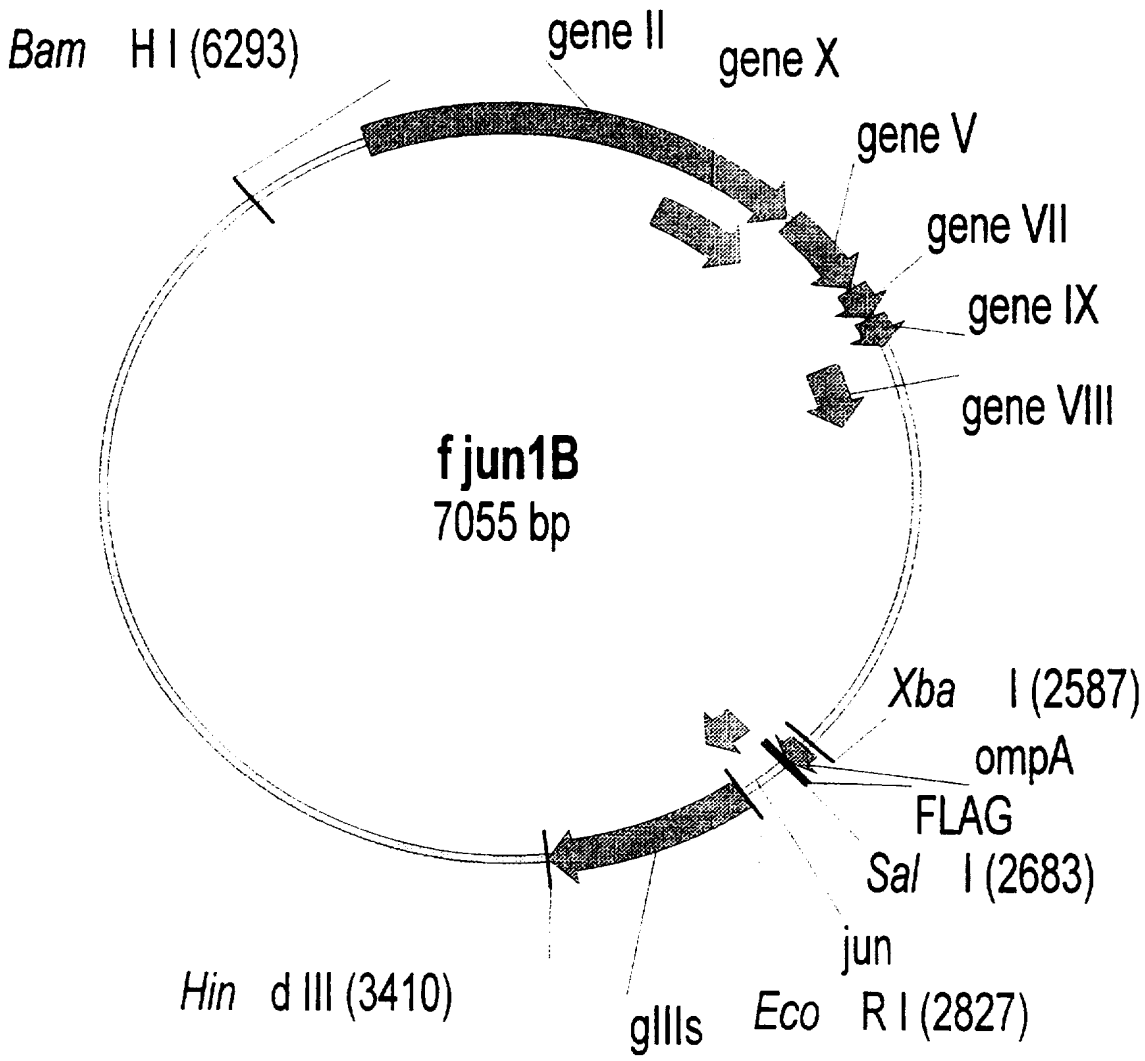


Figure 14: Phage vector fpep3_1B: functional map

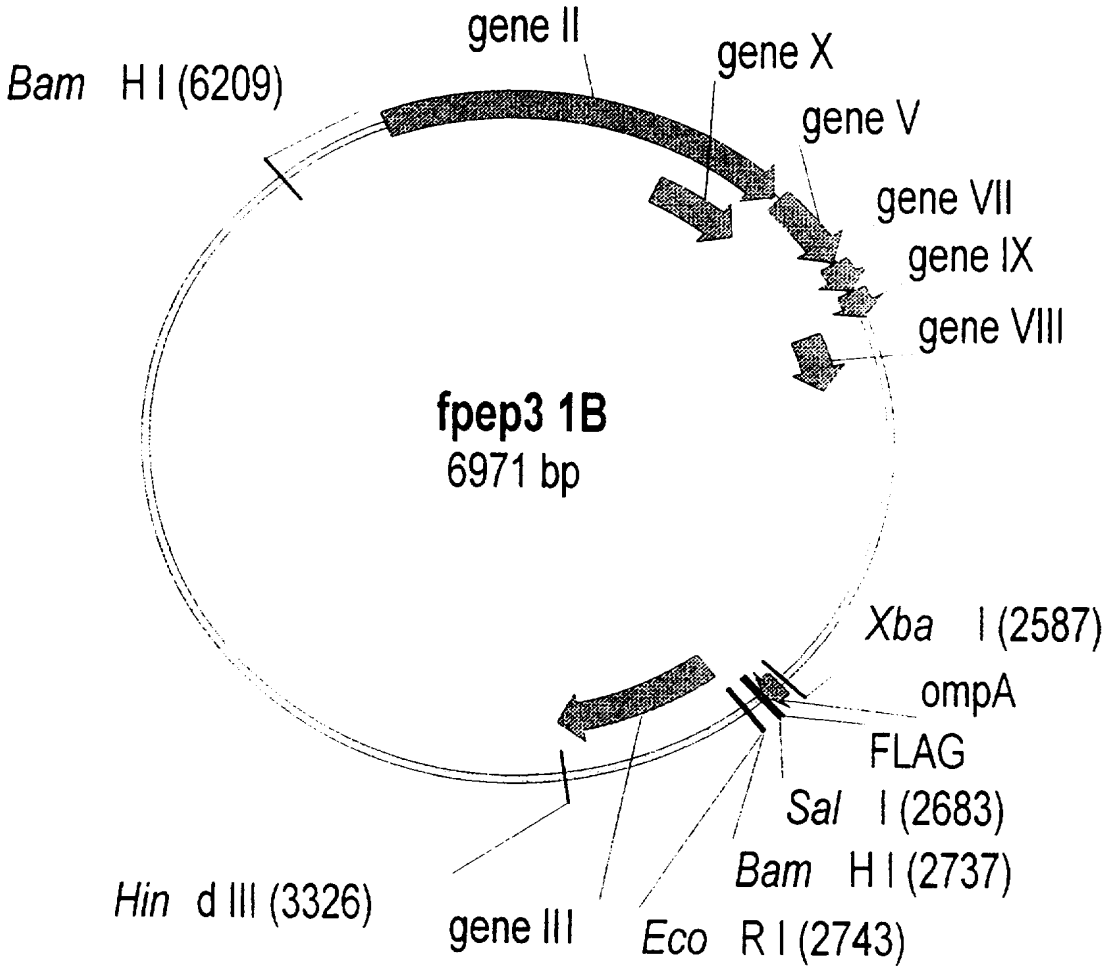


Figure 15: Phage vector fNGF_1B: functional map

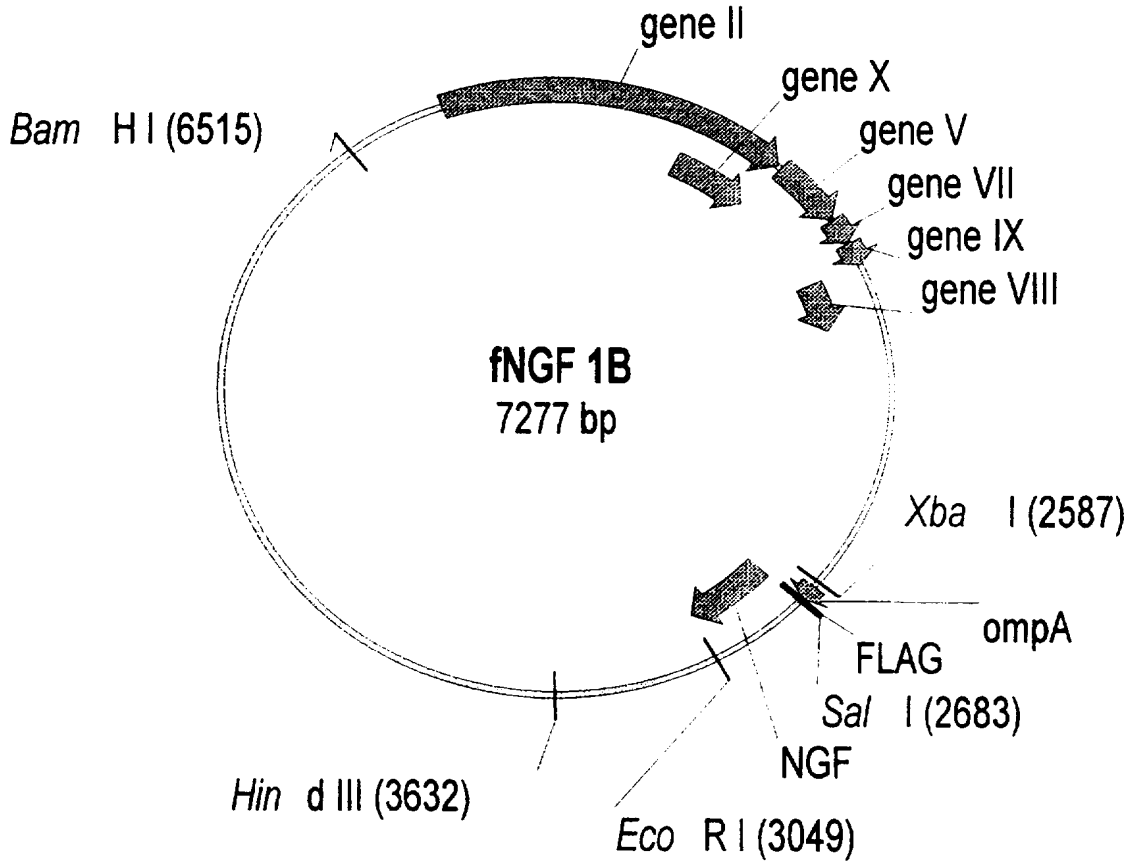


Figure 16: Plasmid pUC19/IMPhag: functional map

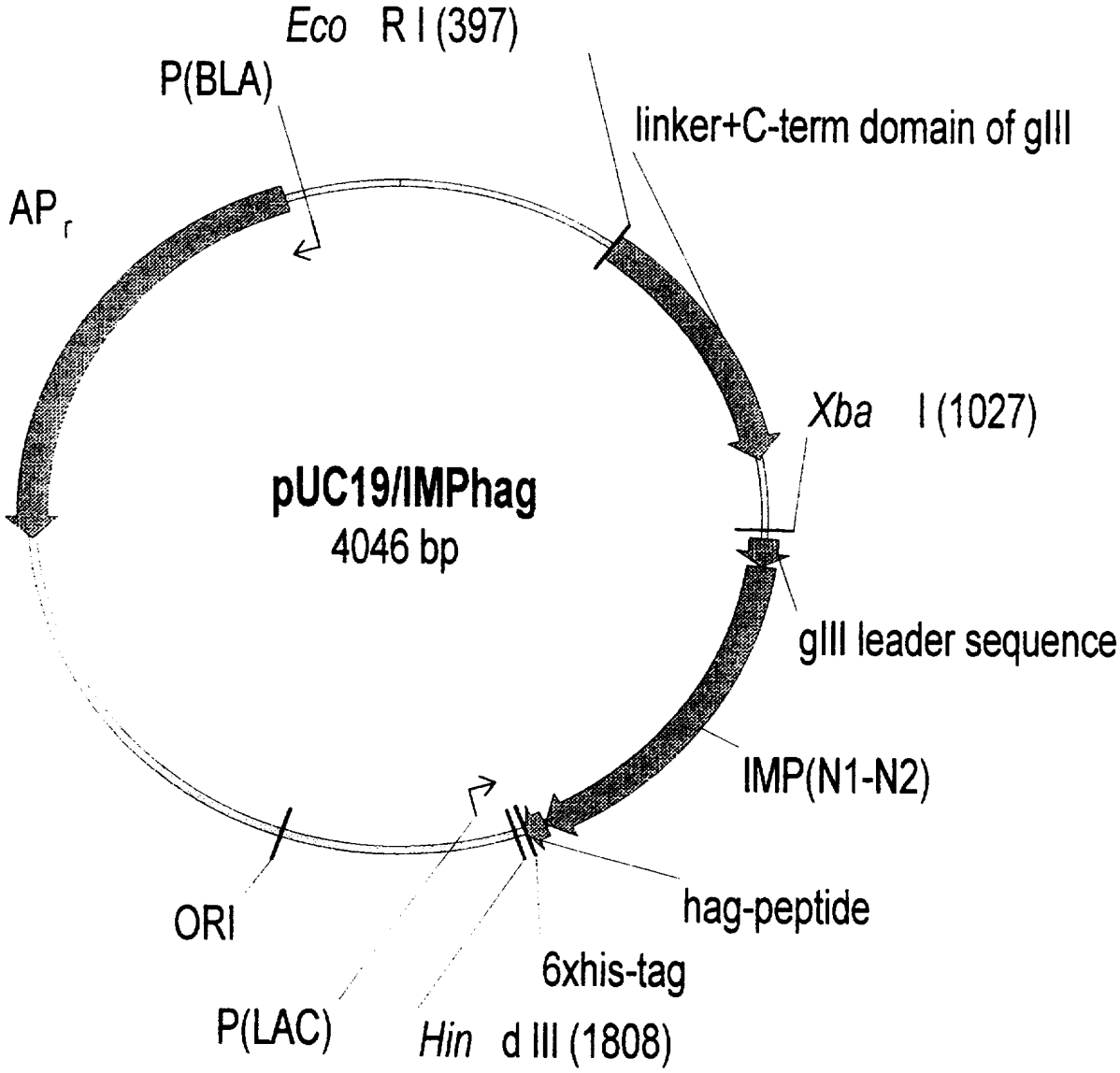


Figure 17: Plasmid pUC18/IMPp75: functional map

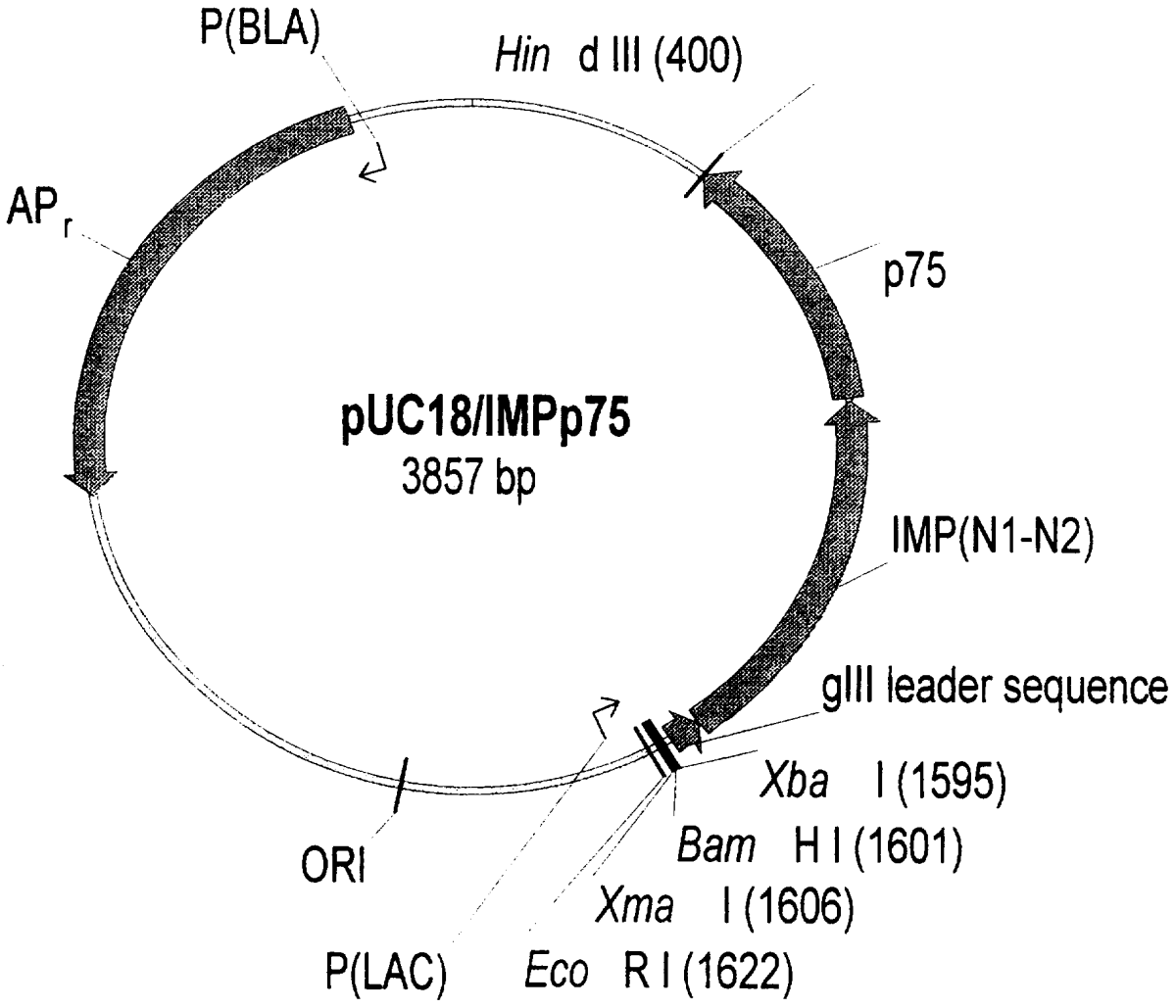


Figure 18: Plasmid pUC18/IMPIL16: functional map

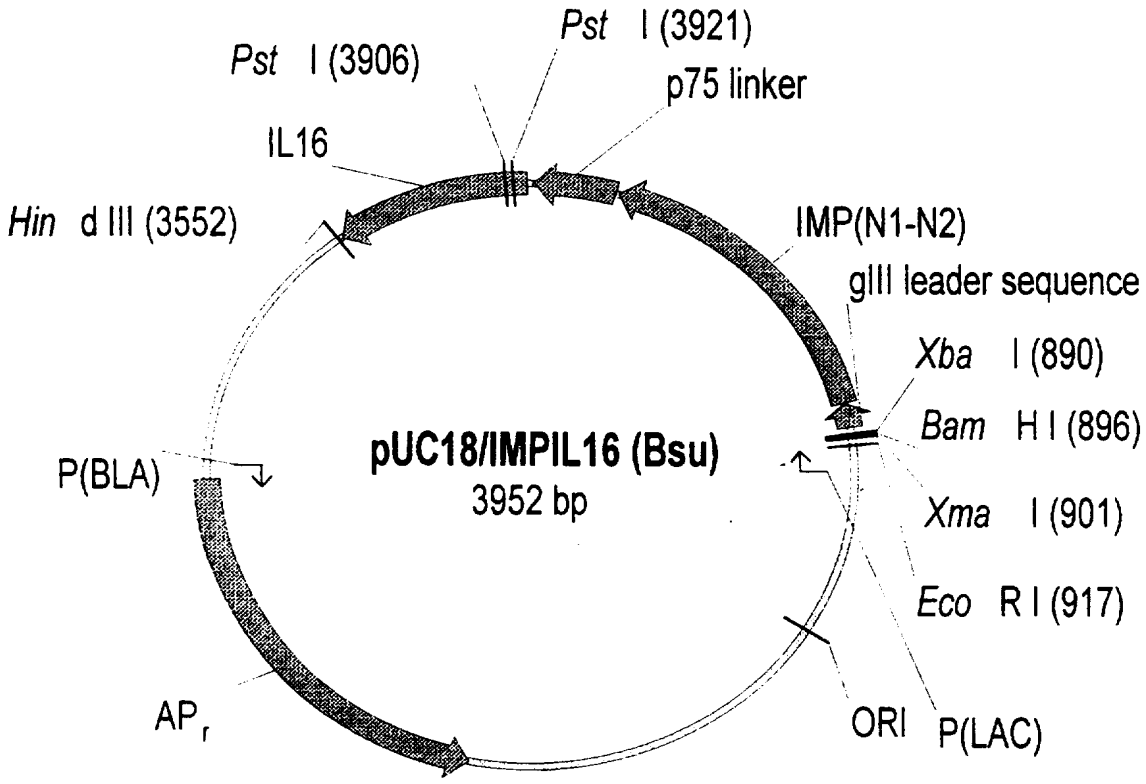


Figure 19: Analysis of selected clones (see Table 3)

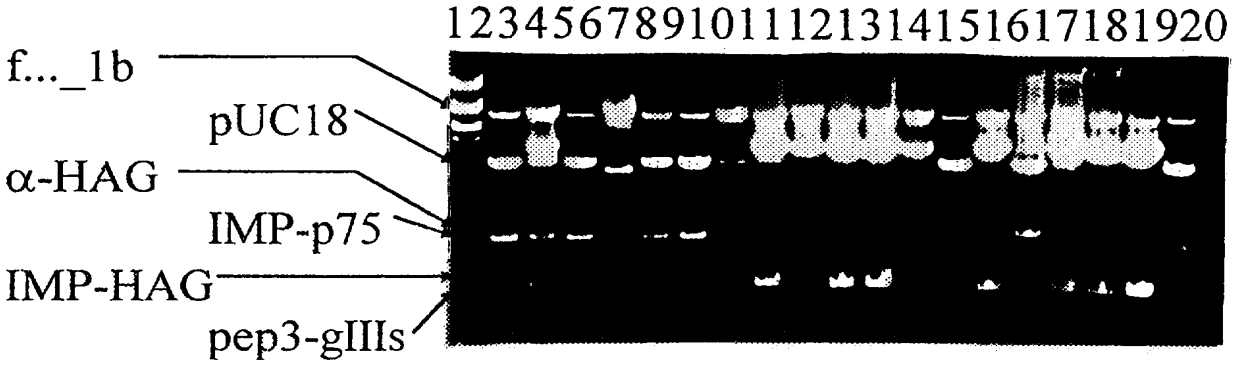


Figure 20: Co-transformation of phagemids, *in vivo* recombination and selection *via* His-tag: general description

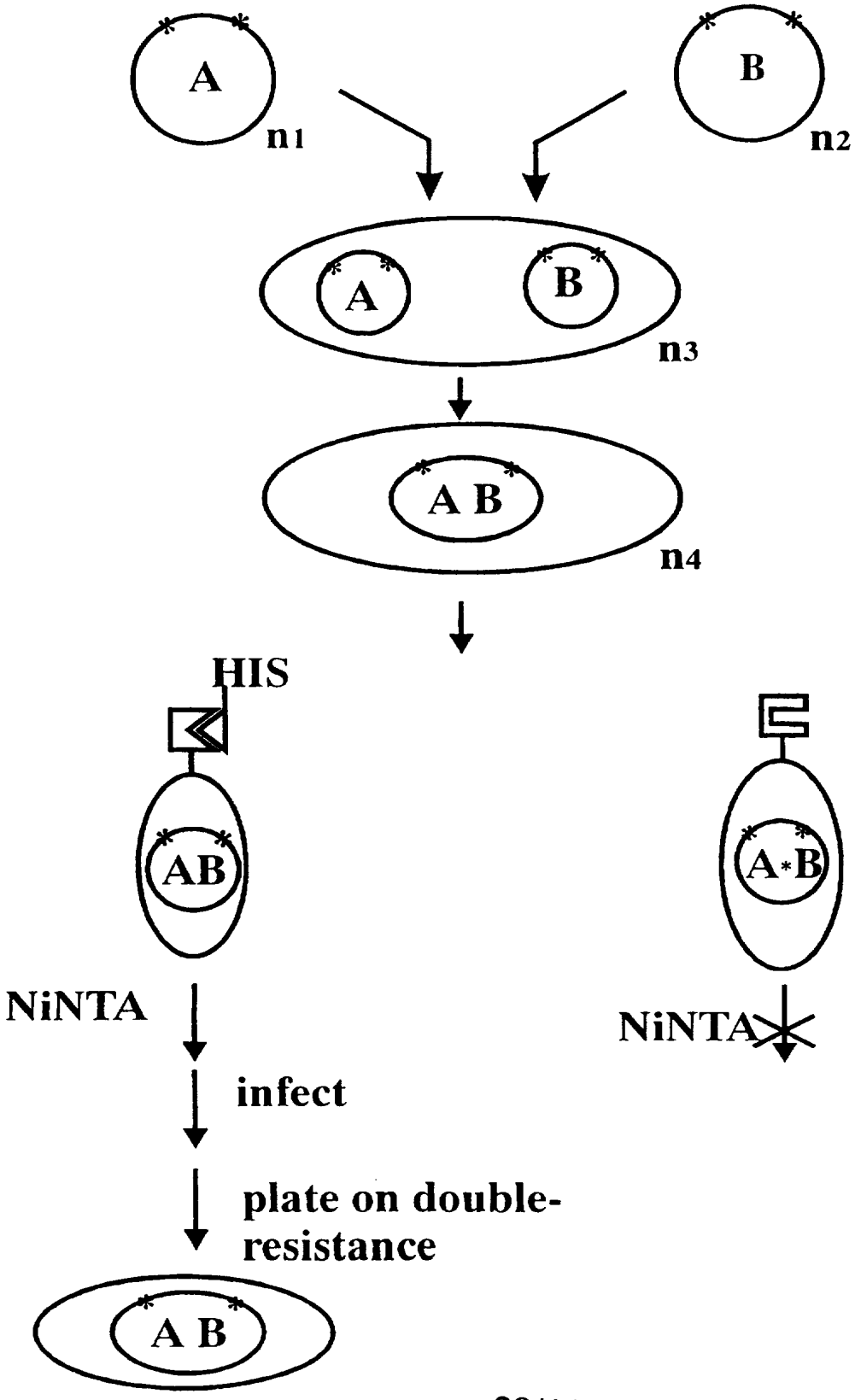


Figure 21: *In vitro* recombination and selection *via* His-tag: general description

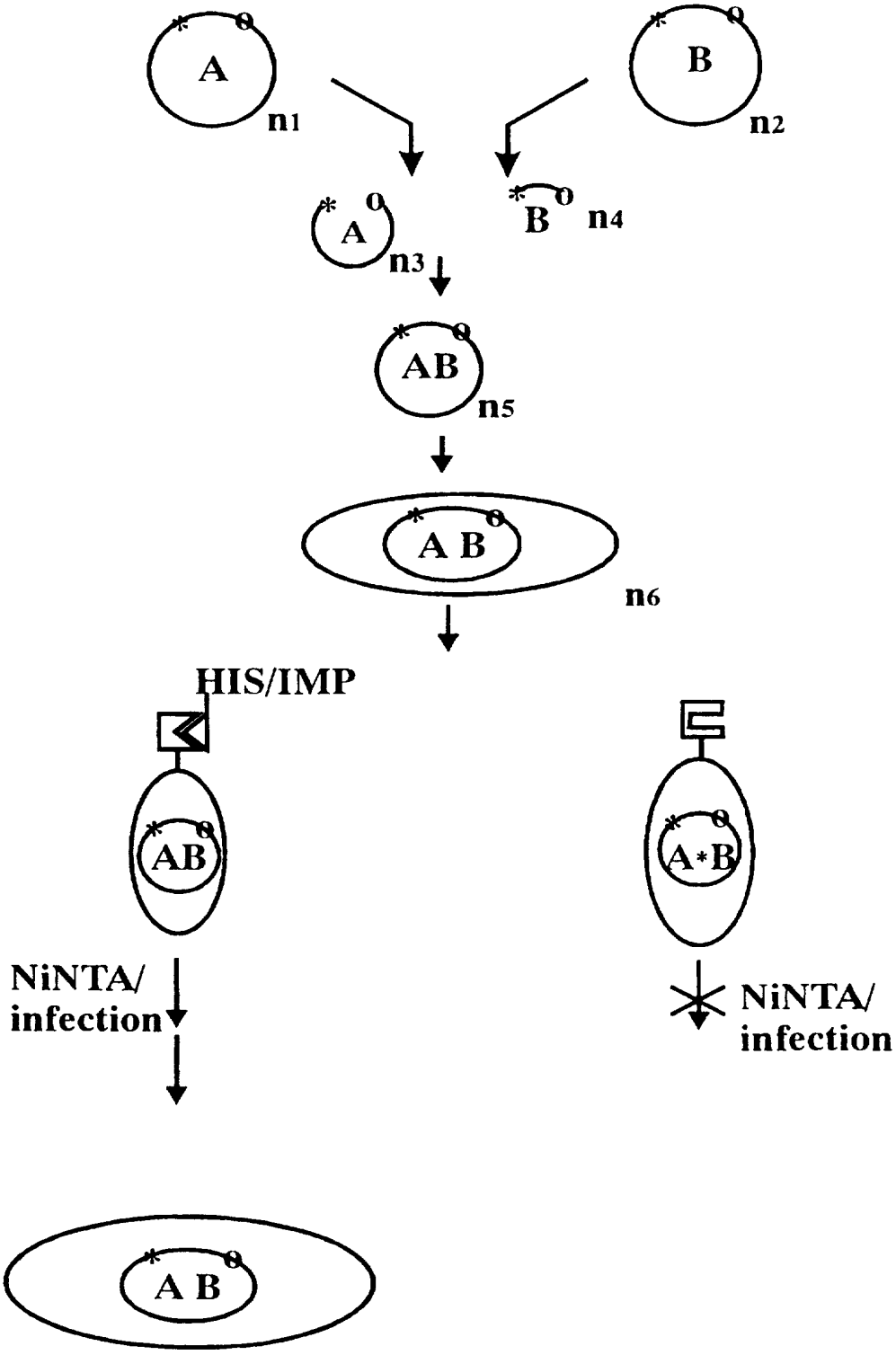


Figure 22: Phage vector fjunhag: functional map

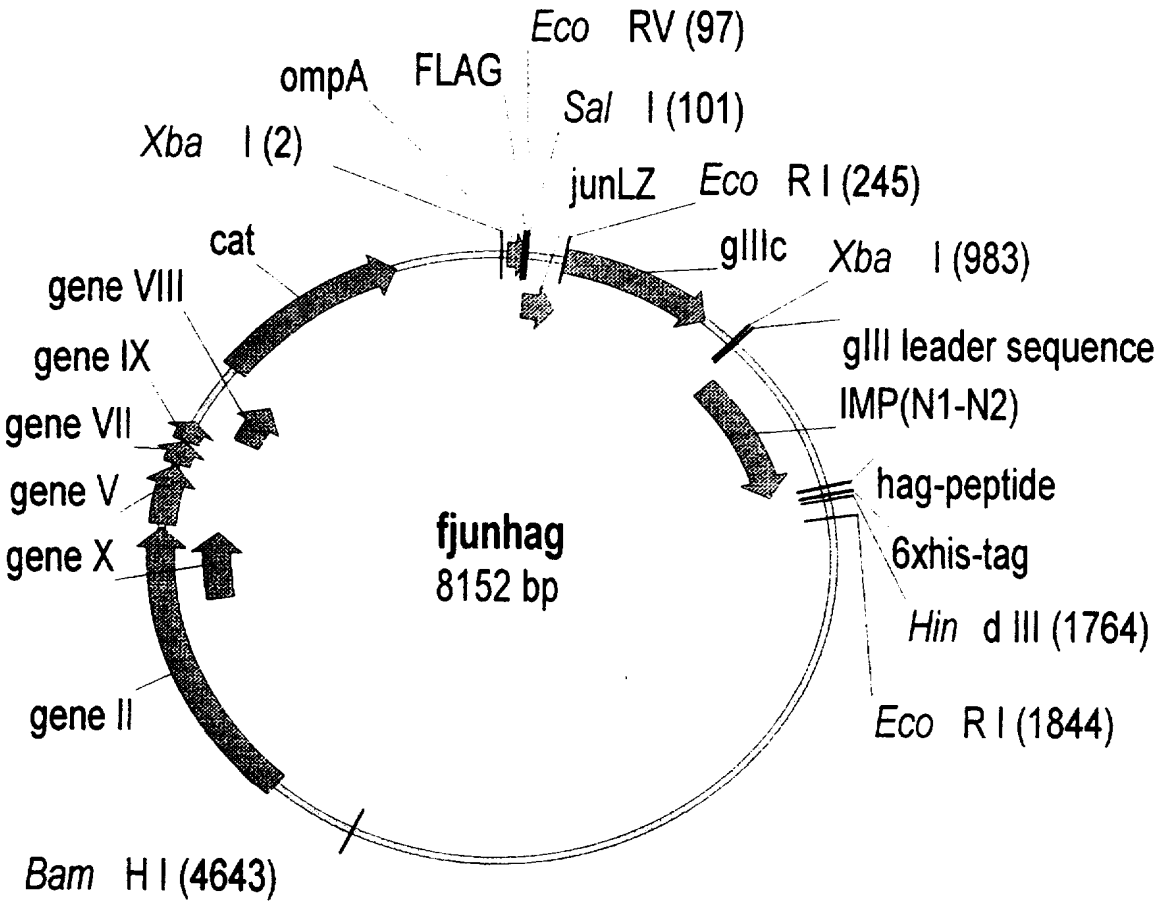


Figure 23: Spatial *in vivo* SIP: general description

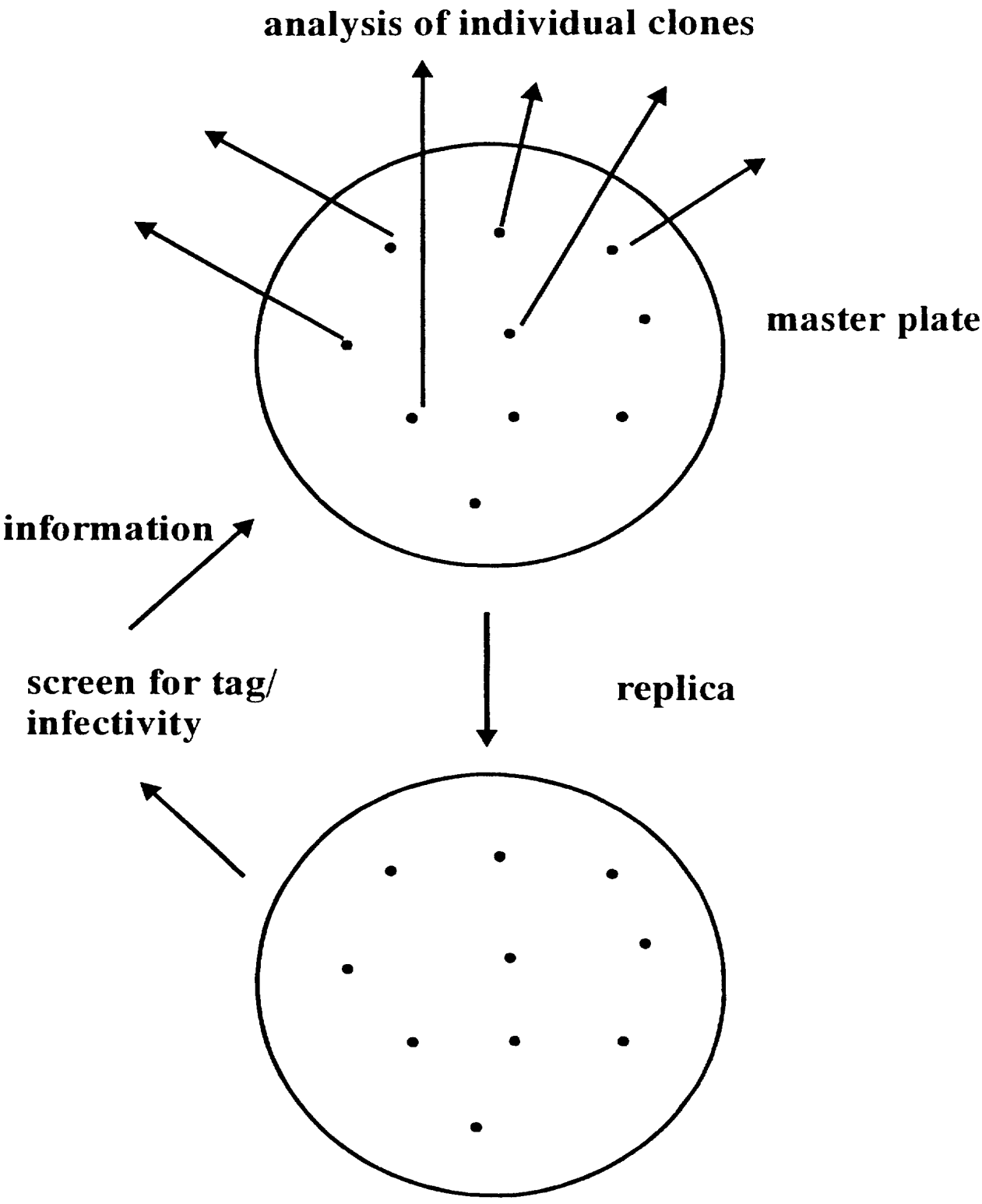


Figure 24: *E. coli* display: general description

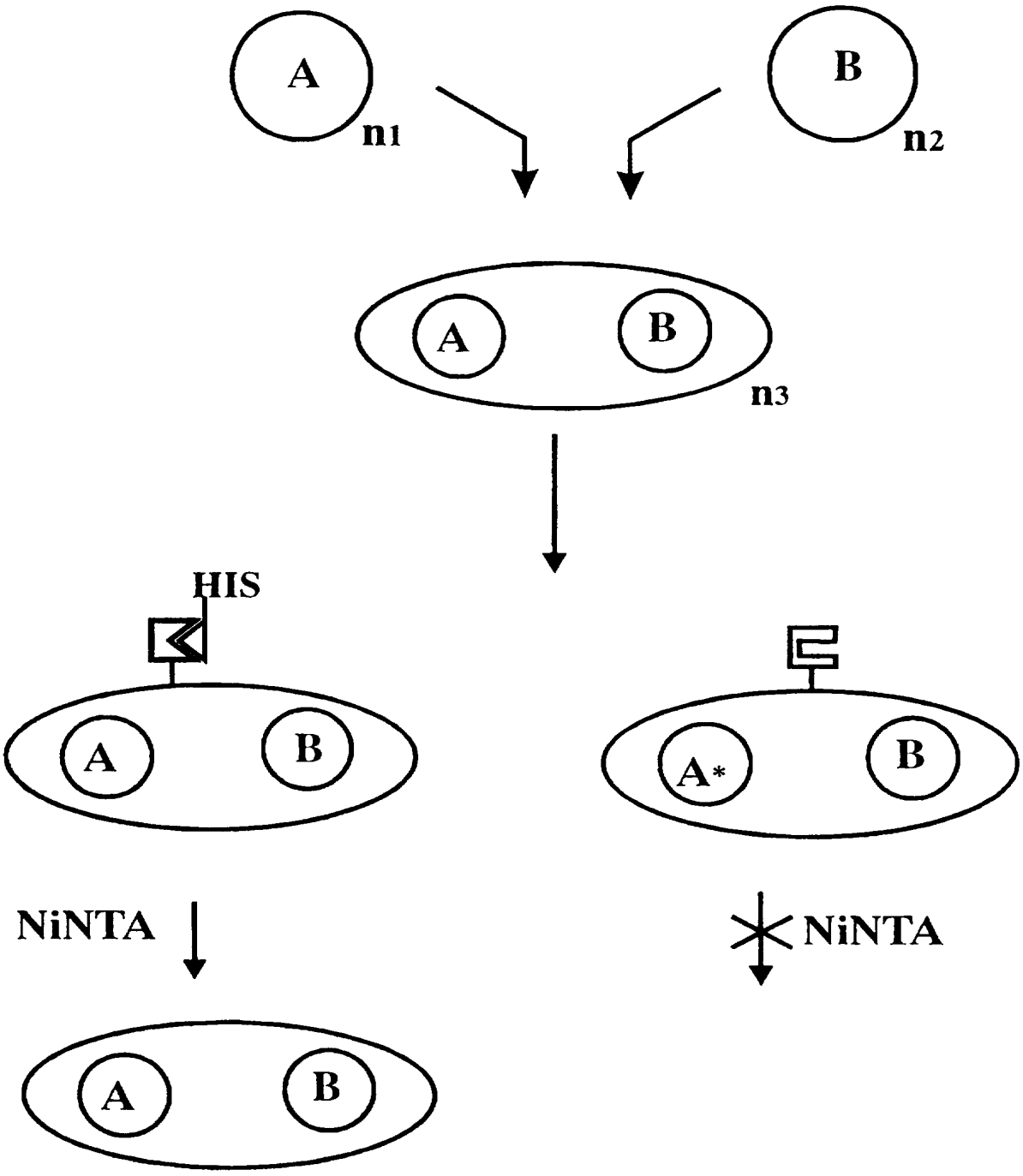


Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence

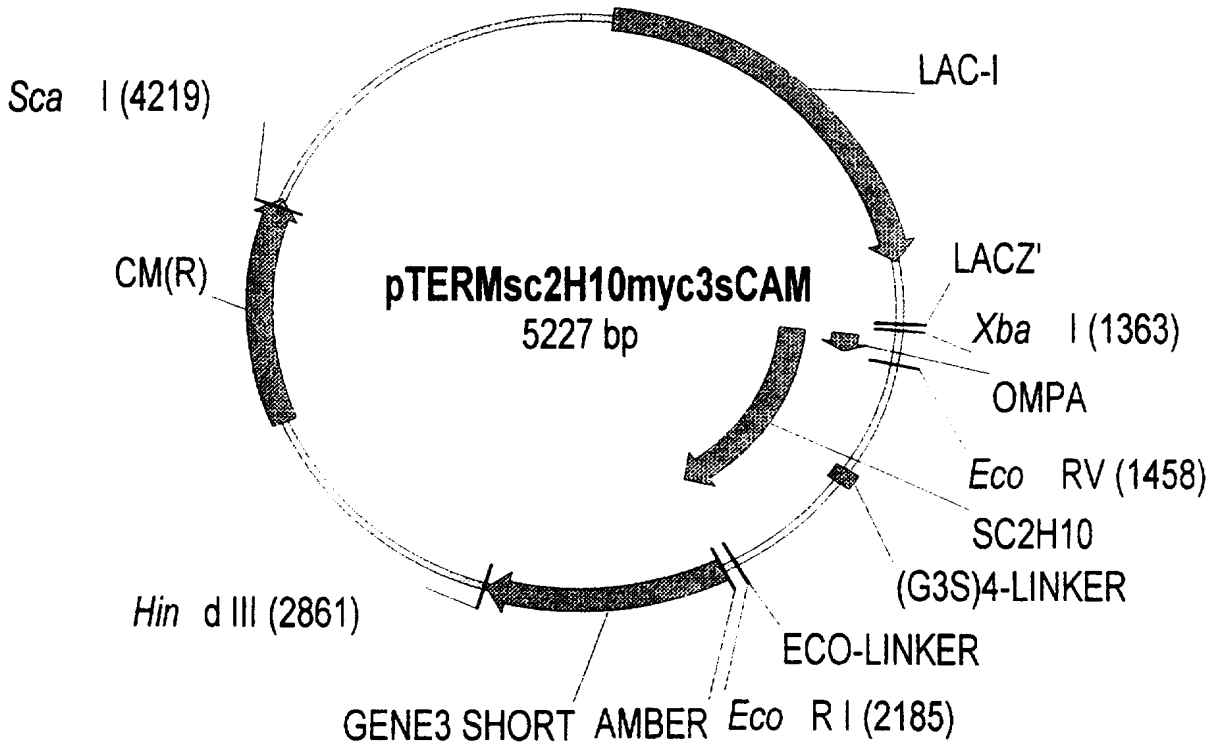


Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
   1  ACCCGACACC ATCGAATGGC GCAAAACCTT TCGCGGTATG GCATGATAGC
      TGGGCTGTGG TAGCTTACCG CGTTTTGGAA AGCGCCATAC CGTACTATCG

  51  GCCCGGAAGA GAGTCAATTC AGGGTGGTGA ATGTGAAACC AGTAACGTTA
      CGGGCCTTCT CTCAGTTAAG TCCCACCACT TACACTTTGG TCATTGCAAT

 101  TACGATGTCG CAGAGTATGC CGGTGTCTCT TATCAGACCG TTTCCCGCGT
      ATGCTACAGC GTCTCATACG GCCACAGAGA ATAGTCTGGC AAAGGGCGCA

 151  GGTGAACCAG GCCAGCCACG TTTCTGCGAA AACGCGGGAA AAAGTGGAAG
      CCACTTGGTC CGGTCGGTGC AAAGACGCTT TTGCGCCCTT TTTCACCTTC

 201  CGGCGATGGC GGAGCTGAAT TACATTCCCA ACCGCGTGGC ACAACAACTG
      GCCGCTACCG CCTCGACTTA ATGTAAGGGT TGGCGCACCG TGTTGTTGAC

 251  GCGGGCAAAC AGTCGTTGCT GATTGGCGTT GCCACCTCCA GTCTGGCCCT
      CGCCCGTTTG TCAGCAACGA CTAACCGCAA CGGTGGAGGT CAGACCGGGA

 301  GCACGCGCCG TCGCAAATTG TCGCGGCGAT TAAATCTCGC GCCGATCAAC
      CGTGCGCGGC AGCGTTTAAC AGCGCCGCTA ATTTAGAGCG CGGCTAGTTG

 351  TGGGTGCCAG CGTGGTGGTG TCGATGGTAG AACGAAGCGG CGTCGAAGCC
      ACCCACGGTC GCACCACCAC AGCTACCATC TTGCTTCGCC GCAGCTTCGG

 401  TGTAAAGCGG CGGTGCACAA TCTTCTCGCG CAACGCGTCA GTGGGCTGAT
      ACATTTCGCC GCCACGTGTT AGAAGAGCGC GTTGCGCAGT CACCCGACTA

 451  CATTAACTAT CCGCTGGATG ACCAGGATGC CATTGCTGTG GAAGCTGCCT
      GTAATTGATA GGCGACCTAC TGGTCCTACG GTAACGACAC CTTCGACGGA

 501  GCACTAATGT TCCGGCGTTA TTTCTTGATG TCTCTGACCA GACACCCATC
      CGTGATTACA AGGCCGCAAT AAAGAACTAC AGAGACTGGT CTGTGGGTAG

 551  AACAGTATTA TTTTCTCCCA TGAAGACGGT ACGCGACTGG GCGTGGAGCA
      TTGTCATAAT AAAAGAGGGT ACTTCTGCCA TGCGCTGACC CGCACCTCGT

 601  TCTGGTCGCA TTGGGTCACC AGCAAATCGC GCTGTTAGCG GGCCCATTAA
      AGACCAGCGT AACCCAGTGG TCGTTTAGCG CGACAATCGC CCGGGTAATT

 651  GTTCTGTCTC GGCGCGTCTG CGTCTGGCTG GCTGGCATAA ATATCTCACT
      CAAGACAGAG CCGCGCAGAC GCAGACCGAC CGACCGTATT TATAGAGTGA

 701  CGCAATCAAA TTCAGCCGAT AGCGGAACGG GAAGGCGACT GGAGTGCCAT
      GCGTTAGTTT AAGTCGGCTA TCGCCTTGCC CTTCCGCTGA CCTCACGGTA

 751  GTCCGGTTTT CAACAAACCA TGCAAATGCT GAATGAGGGC ATCGTTCCCA
      CAGGCCAAAA GTTGTTTGGT ACGTTTACGA CTTACTCCCG TAGCAAGGGT

 801  CTGCGATGCT GGTTGCCAAC GATCAGATGG CGCTGGGCGC AATGCGCGCC
      GACGCTACGA CCAACGGTTG CTAGTCTACC GCGACCCGCG TTACGCGCGG
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
 851  ATTACCGAGT CCGGGCTGCG CGTTGGTGCG GACATCTCGG TAGTGGGATA
      TAATGGCTCA GGCCCGACGC GCAACCACGC CTGTAGAGCC ATCACCCTAT

 901  CGACGATACC GAAGACAGCT CATGTTATAT CCCGCCGTTA ACCACCATCA
      GCTGCTATGG CTTCTGTCGA GTACAATATA GGGCGGCAAT TGGTGGTAGT

 951  AACAGGATTT TCGCCTGCTG GGGCAAACCA GCGTGGACCG CTTGCTGCAA
      TTGTCCTAAA AGCGGACGAC CCCGTTTGGT CGCACCTGGC GAACGACGTT

1001  CTCTCTCAGG GCCAGGCGGT GAAGGGCAAT CAGCTGTTGC CCGTCTCACT
      GAGAGAGTCC CGGTCCGCCA CTTCCCGTTA GTCGACAACG GGCAGAGTGA

1051  GGTGAAAAGA AAAACCACCC TGGCGCCCAA TACGCAAACC GCCTCTCCCC
      CCACTTTTCT TTTTGGTGGG ACCGCGGGTT ATGCGTTTGG CGGAGAGGGG

1101  GCGCGTTGGC CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG
      CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGGCTGAC

1151  GAAAGCGGGC AGTGAGCGGT ACCCGATAAA AGCGGCTTCC TGACAGGAGG
      CTTTCGCCCG TCACTCGCCA TGGGCTATTT TCGCCGAAGG ACTGTCCTCC

1201  CCGTTTTGTT TTGCAGCCCA CCTCAACGCA ATTAATGTGA GTTAGCTCAC
      GGCAAAACAA AACGTCGGGT GGAGTTGCGT TAATTACACT CAATCGAGTG

1251  TCATTAGGCA CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT
      AGTAATCCGT GGGGTCCGAA ATGTGAAATA CGAAGGCCGA GCATACAACA

1301  GTGGAATTGT GAGCGGATAA CAATTTCACA CAGGAAACAG CTATGACCAT
      CACCTTAACA CTCGCCTATT GTTAAAGTGT GTCCTTTGTC GATACTGGTA

                  XbaI
                  ~~~~~~
1351  GATTACGAAT TTCTAGATAA CGAGGGCAAA AAATGAAAAA GACAGCTATC
      CTAATGCTTA AAGATCTATT GCTCCCGTTT TTTACTTTTT CTGTCGATAG

1401  GCGATTGCAG TGGCACTGGC TGGTTTCGCT ACCGTAGCGC AGGCCGACTA
      CGCTAACGTC ACCGTGACCG ACCAAAGCGA TGGCATCGCG TCCGGCTGAT

        EcoRV
        ~~~~~~~
1451  CAAAGATATC GTGATGACCC AGTCTCCAGC AATCATGTCT ACATCTCTAG
      GTTTCTATAG CACTACTGGG TCAGAGGTCG TTAGTACAGA TGTAGAGATC

1501  GGGAACGGGT CACCATGACC TGCACTGCCA GTTCAAGTGT AAGTTCCTCT
      CCCTTGCCCA GTGGTACTGG ACGTGACGGT CAAGTTCACA TTCAAGGAGA

1551  TACTTACACT GGTACCAGCA GAAGCCAGGA TCCTCCCCCA AACTCTGGAT
      ATGAATGTGA CCATGGTCGT CTTCGGTCCT AGGAGGGGGT TTGAGACCTA

1601  TTATAGCACA TCCAACCTGG CTTCTGGAGT CCCAACTCGC TTCAGTGGCA
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
        AATATCGTGT AGGTTGGACC GAAGACCTCA GGGTTGAGCG AAGTCACCGT

1651    GTGGGTCTGG GACCTCTTAC TCTCTCACAA TCAGCACCAT GGCGGCTGAG
        CACCCAGACC CTGGAGAATG AGAGAGTGTT AGTCGTGGTA CCGCCGACTC

1701    GATGCTGCCA CTTATTACTG CCACCAGTAT CATCGTTTCC CACCCACGTT
        CTACGACGGT GAATAATGAC GGTGGTCATA GTAGCAAAGG GTGGGTGCAA

1751    CGGAGGGGGG ACCAAGCTGG AAATAAAACG GGCTGGTGGT GGTGGTTCTG
        GCCTCCCCCC TGGTTCGACC TTTATTTTGC CCGACCACCA CCACCAAGAC

1801    GCGGCGGCGG CTCCGGTGGT GGTGGTTCTG AAGTTAAACT GGTCGAGTCT
        CGCCGCCGCC GAGGCCACCA CCACCAAGAC TTCAATTTGA CCAGCTCAGA

1851    GGAGGAGGCT TGGTGCAACC TGGAGGATCC ATGAAACTCT CCTGTGTTGC
        CCTCCTCCGA ACCACGTTGG ACCTCCTAGG TACTTTGAGA GGACACAACG

1901    CTCTGGAATC ACTTTCAGTA ATTACCGGAT GAACTGGGTC CGCCAGTCTC
        GAGACCTTAG TGAAAGTCAT TAATGGCCTA CTTGACCCAG GCGGTCAGAG

1951    CAGAGAAGGG GCTTGAGTGG GTTGCTGAAA TTAGATTGAA ATCTAATAAT
        GTCTCTTCCC CGAACTCACC CAACGACTTT AATCTAACTT TAGATTATTA

2001    TATGCAACAC ATTATGCGGA GTCTGTGAAA GGGAGGTTCA CCATCTCAAG
        ATACGTTGTG TAATACGCCT CAGACACTTT CCCTCCAAGT GGTAGAGTTC

2051    AGATGATTCC AAAAGTAGTG TCTACCTGCA AATGAACAAC TTAAGAGCTG
        TCTACTAAGG TTTTCATCAC AGATGGACGT TTACTTGTTG AATTCTCGAC

2101    AAGACACTGG CATTTATTAC TGTAGAGGGG TTTCATATAC TATAGACTAC
        TTCTGTGACC GTAAATAATG ACATCTCCCC AAAGTATATG ATATCTGATG

                                                  EcoRI
                                                  ~~~~~~
2151    TGGGGTCAAG GAACCTCAGT CACAGTCTCC TCAGAATTCG AGCAGAAGCT
        ACCCCAGTTC CTTGGAGTCA GTGTCAGAGG AGTCTTAAGC TCGTCTTCGA

2201    GATCTCTGAG GAAGACCTGT AGGCATGCTT ATTTGTTTGT GAATATCAAG
        CTAGAGACTC CTTCTGGACA TCCGTACGAA TAAACAAACA CTTATAGTTC

2251    GCCAATCGTC TGACCTGCCT CAACCTCCTG TCAATGCTGG CGGCGGCTCT
        CGGTTAGCAG ACTGGACGGA GTTGGAGGAC AGTTACGACC GCCGCCGAGA

2301    GGTGGTGGTT CTGGTGGCGG CTCTGAGGGT GGTGGCTCTG AGGGTGGCGG
        CCACCACCAA GACCACCGCC GAGACTCCCA CCACCGAGAC TCCCACCGCC

2351    TTCTGAGGGT GGCGGCTCTG AGGGAGGCGG TTCCGGTGGT GGCTCTGGTT
        AAGACTCCCA CCGCCGAGAC TCCCTCCGCC AAGGCCACCA CCGAGACCAA

2401    CCGGTGATTT TGATTATGAA AAGATGGCAA ACGCTAATAA GGGGGCTATG
        GGCCACTAAA ACTAATACTT TTCTACCGTT TGCGATTATT CCCCCGATAC
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
2451  ACCGAAAATG CCGATGAAAA CGCGCTACAG TCTGACGCTA AAGGCAAACT
      TGGCTTTTAC GGCTACTTTT GCGCGATGTC AGACTGCGAT TTCCGTTTGA

2501  TGATTCTGTC GCTACTGATT ACGGTGCTGC TATCGATGGT TTCATTGGTG
      ACTAAGACAG CGATGACTAA TGCCACGACG ATAGCTACCA AAGTAACCAC

2551  ACGTTTCCGG CCTTGCTAAT GGTAATGGTG CTACTGGTGA TTTTGCTGGC
      TGCAAAGGCC GGAACGATTA CCATTACCAC GATGACCACT AAAACGACCG

2601  TCTAATTCCC AAATGGCTCA AGTCGGTGAC GGTGATAATT CACCTTTAAT
      AGATTAAGGG TTTACCGAGT TCAGCCACTG CCACTATTAA GTGGAAATTA

2651  GAATAATTTC CGTCAATATT TACCTTCCCT CCCTCAATCG GTTGAATGTC
      CTTATTAAAG GCAGTTATAA ATGGAAGGGA GGGAGTTAGC CAACTTACAG

2701  GCCCTTTTGT CTTTGGCGCT GGTAAACCAT ATGAATTTTC TATTGATTGT
      CGGGAAAACA GAAACCGCGA CCATTTGGTA TACTTAAAAG ATAACTAACA

2751  GACAAAATAA ACTTATTCCG TGGTGTCTTT GCGTTTCTTT TATATGTTGC
      CTGTTTTATT TGAATAAGGC ACCACAGAAA CGCAAAGAAA ATATACAACG

2801  CACCTTTATG TATGTATTTT CTACGTTTGC TAACATACTG CGTAATAAGG
      GTGGAAATAC ATACATAAAA GATGCAAACG ATTGTATGAC GCATTATTCC

                HindIII
                ~~~~~~~
2851  AGTCTTGATA AGCTTGACCT GTGAAGTGAA AAATGGCGCA CATTGTGCGA
      TCAGAACTAT TCGAACTGGA CACTTCACTT TTTACCGCGT GTAACACGCT

2901  CATTTTTTTT GTCTGCCGTT TACCGCTACT GCGTCACGGA TCCCCACGCG
      GTAAAAAAAA CAGACGGCAA ATGGCGATGA CGCAGTGCCT AGGGGTGCGC

2951  CCCTGTAGCG GCGCATTAAG CGCGGCGGGT GTGGTGGTTA CGCGCAGCGT
      GGGACATCGC CGCGTAATTC GCGCCGCCCA CACCACCAAT GCGCGTCGCA

3001  GACCGCTACA CTTGCCAGCG CCCTAGCGCC CGCTCCTTTC GCTTTCTTCC
      CTGGCGATGT GAACGGTCGC GGGATCGCGG GCGAGGAAAG CGAAAGAAGG

3051  CTTCCTTTCT CGCCACGTTC GCCGGCTTTC CCCGTCAAGC TCTAAATCGG
      GAAGGAAAGA GCGGTGCAAG CGGCCGAAAG GGGCAGTTCG AGATTTAGCC

3101  GGCATCCCTT TAGGGTTCCG ATTTAGTGCT TTACGGCACC TCGACCCCAA
      CCGTAGGGAA ATCCCAAGGC TAAATCACGA AATGCCGTGG AGCTGGGGTT

3151  AAAACTTGAT TAGGGTGATG GTTCACGTAG TGGGCCATCG CCCTGATAGA
      TTTTGAACTA ATCCCACTAC CAAGTGCATC ACCCGGTAGC GGGACTATCT

3201  CGGTTTTTCG CCCTTTGACG TTGGAGTCCA CGTTCTTTAA TAGTGGACTC
      GCCAAAAAGC GGGAAACTGC AACCTCAGGT GCAAGAAATT ATCACCTGAG
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
3251  TTGTTCCAAA CTGGAACAAC ACTCAACCCT ATCTCGGTCT ATTCTTTTGA
      AACAAGGTTT GACCTTGTTG TGAGTTGGGA TAGAGCCAGA TAAGAAAACT

3301  TTTATAAGGG ATTTTGCCGA TTTCGGCCTA TTGGTTAAAA AATGAGCTGA
      AAATATTCCC TAAAACGGCT AAAGCCGGAT AACCAATTTT TTACTCGACT

3351  TTTAACAAAA ATTTAACGCG AATTTTAACA AAATATTAAC GTTTACAATT
      AAATTGTTTT TAAATTGCGC TTAAAATTGT TTTATAATTG CAAATGTTAA

3401  TCAGGTGGCA CTTTTCGGGG AAATGTGCGC GGAACCCCTA TTTGTTTATT
      AGTCCACCGT GAAAAGCCCC TTTACACGCG CCTTGGGGAT AAACAAATAA

3451  TTTCTAAATA CATTCAAATA TGTATCCGCT CATGTCGAGA CGTTGGGTGA
      AAAGATTTAT GTAAGTTTAT ACATAGGCGA GTACAGCTCT GCAACCCACT

3501  GGTTCCAACT TTCACCATAA TGAAATAAGA TCACTACCGG GCGTATTTTT
      CCAAGGTTGA AAGTGGTATT ACTTTATTCT AGTGATGGCC CGCATAAAAA

3551  TGAGTTATCG AGATTTTCAG GAGCTAAGGA AGCTAAAATG GAGAAAAAAA
      ACTCAATAGC TCTAAAAGTC CTCGATTCCT TCGATTTTAC CTCTTTTTTT

3601  TCACTGGATA TACCACCGTT GATATATCCC AATGGCATCG TAAAGAACAT
      AGTGACCTAT ATGGTGGCAA CTATATAGGG TTACCGTAGC ATTTCTTGTA

3651  TTTGAGGCAT TTCAGTCAGT TGCTCAATGT ACCTATAACC AGACCGTTCA
      AAACTCCGTA AAGTCAGTCA ACGAGTTACA TGGATATTGG TCTGGCAAGT

3701  GCTGGATATT ACGGCCTTTT TAAAGACCGT AAAGAAAAAT AAGCACAAGT
      CGACCTATAA TGCCGGAAAA ATTTCTGGCA TTTCTTTTTA TTCGTGTTCA

3751  TTTATCCGGC CTTTATTCAC ATTCTTGCCC GCCTGATGAA TGCTCATCCG
      AAATAGGCCG GAAATAAGTG TAAGAACGGG CGGACTACTT ACGAGTAGGC

3801  GAGTTCCGTA TGGCAATGAA AGACGGTGAG CTGGTGATAT GGGATAGTGT
      CTCAAGGCAT ACCGTTACTT TCTGCCACTC GACCACTATA CCCTATCACA

3851  TCACCCTTGT TACACCGTTT TCCATGAGCA AACTGAAACG TTTTCATCGC
      AGTGGGAACA ATGTGGCAAA AGGTACTCGT TTGACTTTGC AAAAGTAGCG

3901  TCTGGAGTGA ATACCACGAC GATTTCCGGC AGTTTCTACA CATATATTCG
      AGACCTCACT TATGGTGCTG CTAAAGGCCG TCAAAGATGT GTATATAAGC

3951  CAAGATGTGG CGTGTTACGG TGAAAACCTG GCCTATTTCC CTAAAGGGTT
      GTTCTACACC GCACAATGCC ACTTTTGGAC CGGATAAAGG GATTTCCCAA

4001  TATTGAGAAT ATGTTTTTCG TCTCAGCCAA TCCCTGGGTG AGTTTCACCA
      ATAACTCTTA TACAAAAAGC AGAGTCGGTT AGGGACCCAC TCAAAGTGGT

4051  GTTTTGATTT AAACGTGGCC AATATGGACA ACTTCTTCGC CCCCGTTTTC
      CAAAACTAAA TTTGCACCGG TTATACCTGT TGAAGAAGCG GGGGCAAAAG
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
4101  ACCATGGGCA AATATTATAC GCAAGGCGAC AAGGTGCTGA TGCCGCTGGC
      TGGTACCCGT TTATAATATG CGTTCCGCTG TTCCACGACT ACGGCGACCG

4151  GATTCAGGTT CATCATGCCG TCTGTGATGG CTTCCATGTC GGCAGAATGC
      CTAAGTCCAA GTAGTACGGC AGACACTACC GAAGGTACAG CCGTCTTACG

                          ScaI
                          ~~~~~~~
4201  TTAATGAATT ACAACAGTAC TGCGATGAGT GGCAGGGCGG GGCGTAATTT
      AATTACTTAA TGTTGTCATG ACGCTACTCA CCGTCCCGCC CCGCATTAAA

4251  TTTTAAGGCA GTTATTGGTG CCCTTAAACG CCTGGTGCTA CGCCTGAATA
      AAAATTCCGT CAATAACCAC GGGAATTTGC GGACCACGAT GCGGACTTAT

4301  AGTGATAATA AGCGGATGAA TGGCAGAAAT TCGAAAGCAA ATTCGACCCG
      TCACTATTAT TCGCCTACTT ACCGTCTTTA AGCTTTCGTT TAAGCTGGGC

4351  GTCGTCGGTT CAGGGCAGGG TCGTTAAATA GCCGCTTATG TCTATTGCTG
      CAGCAGCCAA GTCCCGTCCC AGCAATTTAT CGGCGAATAC AGATAACGAC

4401  GTTTACCGGT TTATTGACTA CCGGAAGCAG TGTGACCGTG TGCTTCTCAA
      CAAATGGCCA AATAACTGAT GGCCTTCGTC ACACTGGCAC ACGAAGAGTT

4451  ATGCCTGAGG CCAGTTTGCT CAGGCTCTCC CCGTGGAGGT AATAATTGCT
      TACGGACTCC GGTCAAACGA GTCCGAGAGG GGCACCTCCA TTATTAACGA

4501  CGACATGACC AAAATCCCTT AACGTGAGTT TTCGTTCCAC TGAGCGTCAG
      GCTGTACTGG TTTTAGGGAA TTGCACTCAA AAGCAAGGTG ACTCGCAGTC

4551  ACCCCGTAGA AAAGATCAAA GGATCTTCTT GAGATCCTTT TTTTCTGCGC
      TGGGGCATCT TTTCTAGTTT CCTAGAAGAA CTCTAGGAAA AAAAGACGCG

4601  GTAATCTGCT GCTTGCAAAC AAAAAAACCA CCGCTACCAG CGGTGGTTTG
      CATTAGACGA CGAACGTTTG TTTTTTTGGT GGCGATGGTC GCCACCAAAC

4651  TTTGCCGGAT CAAGAGCTAC CAACTCTTTT TCCGAAGGTA ACTGGCTTCA
      AAACGGCCTA GTTCTCGATG GTTGAGAAAA AGGCTTCCAT TGACCGAAGT

4701  GCAGAGCGCA GATACCAAAT ACTGTCCTTC TAGTGTAGCC GTAGTTAGGC
      CGTCTCGCGT CTATGGTTTA TGACAGGAAG ATCACATCGG CATCAATCCG

4751  CACCACTTCA AGAACTCTGT AGCACCGCCT ACATACCTCG CTCTGCTAAT
      GTGGTGAAGT TCTTGAGACA TCGTGGCGGA TGTATGGAGC GAGACGATTA

4801  CCTGTTACCA GTGGCTGCTG CCAGTGGCGA TAAGTCGTGT CTTACCGGGT
      GGACAATGGT CACCGACGAC GGTCACCGCT ATTCAGCACA GAATGGCCCA

4851  TGGACTCAAG ACGATAGTTA CCGGATAAGG CGCAGCGGTC GGGCTGAACG
      ACCTGAGTTC TGCTATCAAT GGCCTATTCC GCGTCGCCAG CCCGACTTGC

4901  GGGGGTTCGT GCACACAGCC CAGCTTGGAG CGAACGACCT ACACCGAACT
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
        CCCCCAAGCA CGTGTGTCGG GTCGAACCTC GCTTGCTGGA TGTGGCTTGA

4951    GAGATACCTA CAGCGTGAGC TATGAGAAAG CGCCACGCTT CCCGAAGGGA
        CTCTATGGAT GTCGCACTCG ATACTCTTTC GCGGTGCGAA GGGCTTCCCT

5001    GAAAGGCGGA CAGGTATCCG GTAAGCGGCA GGGTCGGAAC AGGAGAGCGC
        CTTTCCGCCT GTCCATAGGC CATTCGCCGT CCCAGCCTTG TCCTCTCGCG

5051    ACGAGGGAGC TTCCAGGGGG AAACGCCTGG TATCTTTATA GTCCTGTCGG
        TGCTCCCTCG AAGGTCCCCC TTTGCGGACC ATAGAAATAT CAGGACAGCC

5101    GTTTCGCCAC CTCTGACTTG AGCGTCGATT TTTGTGATGC TCGTCAGGGG
        CAAAGCGGTG GAGACTGAAC TCGCAGCTAA AAACACTACG AGCAGTCCCC

5151    GGCGGAGCCT ATGGAAAAAC GCCAGCAACG CGGCCTTTTT ACGGTTCCTG
        CCGCCTCGGA TACCTTTTTG CGGTCGTTGC GCCGGAAAAA TGCCAAGGAC

5201    GCCTTTTGCT GGCCTTTTGC TCACATG
        CGGAAAACGA CCGGAAAACG AGTGTAC
```

Table 1: Phagemids Constructed for Experiments 2 and 3

Name	FLAG	His6	gIII	Size (bp)	Insert	REN1	REN2	Resistance
pING1-A1	-	+	-	3783	His	EcoRV	SmaI	Ap
pING1-A2	-	-	-	3795	Strep-tag	EcoRV	SmaI	Ap
pING3-A1	+	+	-	3792	His	EcoRV	SmaI	Ap
pING3-A2	+	-	-	3804	Strep-tag	EcoRV	SmaI	Ap
pONG3-A	+	-	+	4278	-	EcoRV	SmaI	Ap
pYANG3-A	+	-	+	4404	Jun	EcoRV	EcoRI	Ap
pYANG3-Ape2	+	-	+		pep2	XbaI	HindIII	Ap
pYANG3-Ape3	+	-	+		pep3	XbaI	HindIII	Ap
pYANG3-Ape10	+	-	+		pep10	XbaI	HindIII	Ap
pING1-C1	-	+	-	3853	His	EcoRV	SmaI	Cm
pING1-C2	-	-	-	3865	Strep-tag	EcoRV	SmaI	Cm
pING3-C1	+	+	-	3862	His	EcoRV	SmaI	Cm
pING3-C2	+	-	-	3874	Strep-tag	EcoRV	SmaI	Cm
pYING3-C1	+	+	-	3994	Fos	EcoRV	EcoRI	Cm
pYING3-C2	+	+	-	4315	p75	EcoRV	EcoRI	Cm
pYING3-C3	+	+	-	4240	IL-16	EcoRV	EcoRI	Cm

Table 2: Results of Experiment 2 (see Figure 7)

Table 2a: Combination of phagemids present in initial library (α)

	Combination	Clone(s)
1.	pYING1-C2 + pYANG3-ApeX	6
2.	pYING1-C1 + pYANG3-A	1
3.	pYING1-C1 + pYANG3-ApeX	1
4.	pYING1-C2 + pYANG3-A	1
5.	pYING1-C2 + ?	1

Table 2b: Combination of phagemids present after selection (β)

	Combination	Clone(s)
1.	pYING1-C2 + pYANG3-ApeX	1
2.	pYING1-C1 + pYANG3-A	9

Table 3: Results of Experiment 4 (see Figure 19)

Table 3a: Identification of phage/plasmid present in individual clones

Combination	Clone(s)
fhag1A + pUC19/IMPhag	#9
fpep3_1b + pUC18/IMP-p75	#1,#3,#5,#6,#7,#13,#15,#19
fpep3_1b + pUC19/IMPhag	#14
unusual DNA	#2,#4,#8,#10,#11,#12,#16,#17,#18

Table 3b: Test for infectivity of individual clones

Clone #	Titer (transducing units/ml)
1	2 x 10E4
2	31
3	1 x 10E5
4	1 x 10E5
5	1 x 10E5
6	2 x 10E3
7	1 x 10E4
8	1 x 10E5
9	1 x 10E6
10	1 x 10E4
11	1 x 10E3
12	1 x 10E4
13	3 x 10E3
14	< 10
15	5 x 10E4
16	1 x 10E4
17	5 x 10E2
18	1 x 10E4
19	1 x 10E5

INTERNATIONAL SEARCH REPORT

Intern al Application No
PCT/EP 97/00931

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/10 C12N15/12 C12N15/24 C12N15/62 C12N15/70

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 93 19172 A (CAMBRIDGE ANTIBODY TECH ;MEDICAL RES COUNCIL (GB)) 30 September 1993 see the whole document ---	1-7,11, 12, 19-27, 29,30,32
A	EP 0 614 989 A (MORPHOSYS GES FUER PROTEINOPTI) 14 September 1994 cited in the application see the whole document ---	1-34
A	DE 41 22 598 C (DKFZ) 30 July 1992 see the whole document ---	1-34
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 June 1997	03.07.97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	HORNIG, H

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Interr. al Application No
PCT/EP 97/00931

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GENE, vol. 137, 1993, ELSEVIER SCIENCE PUBLISHERS,B.V.,AMSTERDAM,NL;, pages 69-75, XP002033343 R. CAMERI AND M. SUTER: "Display of biological active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to genetic information responsible for their production" cited in the application see the whole document ---	1-34
A	NUCLEIC ACIDS RESEARCH, vol. 22, no. 25, 1994, IRL PRESS LIMITED,OXFORD,ENGLAND, pages 5761-5762, XP002033344 K. GRAMATIKOFF ET AL.: "Direct interaction rescue, a novel filamentous phage technique to study protein - protein interactions" cited in the application see the whole document ---	1-34
A	NATURE BIOTECHNOLOGY, vol. 13, April 1995, NATURE PUBL. CO.,NEW YORK, US, pages 373-377, XP002033345 D. NERI ET AL.: "Calmodulin as a versatile tag for antibody fragments" see the whole document ---	1-34
A	NATURE BIOTECHNOLOGY, vol. 13, April 1995, NATURE PUBL. CO.,NEW YORK, US, pages 366-372, XP002033346 Z. LU ET AL.: "Expression of thioredoxin random peptide libraries on the Escherichia coli cell surface as functional fusions to flagellin: A system designed for exploring protein-protein interactions" cited in the application see the whole document ---	1-34
A	FEBS LETTERS, vol. 377, no. 2, 18 December 1995, ELSEVIER, AMSTERDAM, NL, pages 227-231, XP002033347 C. KREBBER ET AL.: "Co-selection of cognate antibody-antigen pairs by selectively-infective phage" cited in the application see the whole document --- -/--	1-34

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte ›nal Application No
PCT/EP 97/00931

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	WO 97 10330 A (BIOINVENT INT AB ;KIDDLE SIMON JOHN (GB); BORREBAECK CARL (SE); MA) 20 March 1997 see the whole document	1,2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

...ormation on patent family members

Intern ...l Application No
PCT/EP 97/00931

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9319172 A	30-09-93	AU 665190 B	21-12-95
		AU 1693892 A	30-12-92
		AU 665025 B	14-12-95
		AU 2593392 A	27-04-93
		AU 665221 B	21-12-95
		AU 3089092 A	28-06-93
		AU 673515 B	14-11-96
		AU 3763893 A	21-10-93
		CA 2109602 A	26-11-92
		CA 2119930 A	01-04-93
		CA 2124460 A	10-06-93
		CA 2131151 A	30-09-94
		EP 0585287 A	09-03-94
		EP 0605522 A	13-07-94
		EP 0616640 A	28-09-94
		EP 0656941 A	14-06-95
		WO 9220791 A	26-11-92
		WO 9306213 A	01-04-93
		WO 9311236 A	10-06-93
		JP 6510671 T	01-12-94
		JP 6508511 T	29-09-94
		JP 7502167 T	09-03-95
		JP 7505055 T	08-06-95
		US 5565332 A	15-10-96
EP 0614989 A	14-09-94	CA 2115811 A	18-08-94
		DE 614989 T	28-09-95
		JP 7039381 A	10-02-95
		US 5514548 A	07-05-96
DE 4122598 C	30-07-92	WO 9301287 A	21-01-93
		EP 0547200 A	23-06-93
		JP 6501395 T	17-02-94
		US 5591604 A	07-01-97
WO 9710330 A	20-03-97	AU 6939196 A	01-04-97